Stefan Heinz

# Mathematical Modeling

# Mathematical Modeling

Stefan Heinz

# Mathematical Modeling

Springer

Prof. Dr. Stefan Heinz
Department of Mathematics
University of Wyoming
E. University Ave. 1000
82071 Laramie Wyoming
USA
Heinz@uwyo.edu

ISBN 978-3-642-20310-7 e-ISBN 978-3-642-20311-4
DOI 10.1007/978-3-642-20311-4
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011932690



*Cover design*: SPi Publisher Services

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To my wife Petra,
and to our children Josephine and Jakob

# Preface

This book is intended as a text on mathematical modeling for undergraduate and graduate students of mathematics, engineering, economics, finance, biology, chemistry, and physics. The material follows the author's undergraduate teaching of Introduction to Mathematical Modeling and graduate teaching of Deterministic and Stochastic Mathematical Modeling over the last ten years. The first characteristic feature of this text is the systematic development of deterministic and stochastic modeling approaches. Such a development is relevant because most real-world processes involve randomness. The consideration of stochastic methods enables a comprehensive understanding, for example, of the basis of optimal deterministic models and how closed deterministic equations can be obtained. The second characteristic feature of this text is the systematic discussion of single problems: the analysis of observations, characteristic properties and changes of one variable, and the laws that govern the evolution of one and several variables. An alternative approach would be the simultaneous discussion of difference and differential equations, or the simultaneous discussion of differential equations for one and several variables. The latter approach would make the presentation more difficult to understand because of the simultaneous explanation of the laws of stochastic evolution and meaning of stochastic concepts, or the explanation of the laws of stochastic evolution in notation for several variables. The third characteristic feature of this text is a hierarchical development of models (if possible). Examples for this approach are the discussion of statistically most-likely probability density functions, the relations between difference and differential equations, the Brownian motion model and diffusion model, the delay logistic model, non-Markovian and Markovian velocity models, the nonlinear and linear pendulum motion, and the derivation of equations for fluid dynamics in Chaps. 4–10, respectively. Such a systematic discussion of models is relevant to see the usual hierarchical structure of models and the range of applicability of certain models.

A fourth characteristic feature of this text is the attempt to provide a complete presentation. Most of the derivations are presented by providing all the required details (exercises are used to provide additional details). Discussions of problems are provided as complete as possible (see the discussion of Lorenz's equations).

The book enables the illustration of the application of the developed concepts by 570 exercise questions organized in 220 problems. The detailed solutions to all questions are given in the Instructor's Solutions Manual, which can be provided to instructors by the publisher. The exercises are given at the end of each chapter. For example, the notation "Exercise 4.2.4" refers to the fourth problem related to Sect. 4.2. The exercises provided here are much more than sufficient for the assignment of homework. In many cases, it is possible to create additional homework by minor modifications of the suggested problems.

Apart from basic algebra, a first essential prerequisite for following the material presented in this book is single-variable calculus (I and II): students have to be familiar with differentiation and integration, the calculation of local extreme values, and the Taylor series. Complex numbers will be used in Chaps. 5, 7, and 9. Prior knowledge of differential equations is helpful for understanding Chaps. 7–10, but it is not a requirement. The solution of separable differential equations, which is the usual type of equation involved, is explained in detail in Chap. 7. Chapter 9 involves the use of two-by-two and three-by-three matrices. Chapter 10 applies multivariate calculus concepts, e.g., integrals over functions of several variables, partial derivatives, and partial differential equations. However, these developments will be shown to represent simple extensions of single-variable calculus and related methods (for the solution of the Fokker-Planck equation).

**Table 1** Overview of chapters and questions addressed in this book.

| | |
|---|---|
| **1. Determ. Analysis Observations**<br>*How can we develop models that describe the trend of observations?* | **2. Stoch. Analysis Observations**<br>*How can we find optimal models that account for data randomness?* |
| **3. Determ. States**<br>*What are characteristic properties of a deterministic variable?* | **4. Stoch. States**<br>*What are characteristic properties of a random variable?* |
| **5. Determ. Changes**<br>*What are characteristic changes of a deterministic variable?* | **6. Stoch. Changes**<br>*What are characteristic changes of a random variable?* |
| **7. Determ. Evolution**<br>*What are the laws of the evolution of one deterministic variable?* | **8. Stoch. Evolution**<br>*What are the laws of the evolution of one random variable?* |
| **9. Determ. Multivariate Evolution**<br>*What are the laws of the evolution of several deterministic variables?* | **10. Stoch. Multivariate Evolution**<br>*What are the laws of the evolution of several random variables?* |

A second essential prerequisite for applying the concepts developed in this text is basic knowledge of a software package (like Matlab, Mathematica, Maple) that can be used to perform relatively simple numerical calculations. Students should be able to read in given data, apply simple data transformations, analyze the data, and show the results of computations in figures. Examples for calculations that students should be able to perform after the explanation of the corresponding concepts are the plot of model functions in comparison to random data, the calculation of a probability density function, and the numerical solution of ordinary differential equations. From the author's view point, such numerical exercises do essentially contribute to the understanding of students. In fact, only the exercises of Chaps. 1, 2, 4, and 6 (and two exercises of Chap. 9) require the use software.

A first possibility of using this text is to apply the first four chapters (the upper box in Table 1) for the teaching of a three-credit undergraduate course Introduction to Mathematical Modeling. This course would be focused on a basic understanding of how simple analytical functions can be used for the modeling of many real-world problems. The students learn to describe the trend of observations, to deal with the need to consider several variables, to design optimal models, and to assess randomness. The level of this course would be comparable to Calculus II. Instead of covering all the sections of Chap. 4, it is a good alternative to focus on models for probability density functions (Sect. 4.3) combined with Sect. 10.2, which explains (in difference to the approach used in Chap. 2) the development of optimal models from a probability perspective. Another possibility is given by covering Chap. 5 instead of Chap. 4.

A second possibility of using this text is to apply Chaps. 5, 7, and 9 (the lower left-hand side box in Table 1) for the teaching of a three-credit undergraduate and graduate course Deterministic Mathematical Modeling. Such a course would be focused on a basic understanding of how real-world processes can be modeled on the basis of deterministic ordinary differential equations. The students learn about the application, typical advantages and disadvantages of difference and differential equations, the characteristic changes and evolution laws of deterministic processes, and the ways to model and analyze the interaction of processes. The level of this course would be comparable to Applied Differential Equations I and II courses. The difference to usual Applied Differential Equations courses would be the clear focus on the application of typical differential equations. The discussion of Chaps. 5, 7, and 9 combined with additional practice problems taken from the exercises (or taken from Haberman 1977, Fulford et al. 1997, Edelstein-Keshet 2005, Brannan & Boyce 2007, Nagle et al. 2008, Boyce & DiPrima 2009) does provide sufficient material for a one-semester course. A possibility to provide a broader perspective would be the additional discussion of Sects. 3.2 and 3.3 related to the application of dimensional analysis in the beginning of this course.

A third possibility of using this text is to apply Chaps. 6, 8, and 10 (the lower right-hand side box in Table 1) for the teaching of a three-credit graduate course Stochastic Mathematical Modeling. Required knowledge about basic properties of random variables can be provided by involving a part of Chap. 4 (Sects. 4.2 – 4.4). Such a course would be focused on a basic understanding of how random real-world processes can be analyzed and modeled on the basis of stochastic ordinary differential equations. The students learn about the relationship between stochastic differential equations and evolution equations for probability density functions, the characteristic features of Monte Carlo simulation, the way to develop stochastic models, and the use of stochastic methods for developing consistent models for multi-scale processes. Because of the inclusion of stochastic methods, ordinary and partial differential equations, the level of this course would be higher than the levels of the two courses described above. A way to exclude multivariate calculus concepts is to focus the presentation on Chaps. 4, 6, 8, and Sects. 10.2 and 10.3. There are many ways to illustrate the use of stochastic methods in particular applications, for example with regard to financial mathematics (Buchanan 2008), biology (Allen 2003), turbulent reacting flows (Haworth 2010), flow in porous media (Tyagi et al. 2008), two-phase flows (Minier & Peirano 2001), or many other applications (Kloeden & Platen 1992).

It is a pleasure to thank many people for significant support over many years: In particular, I am profoundly grateful to the Professors F. Jafari, P. Jenny, P. Givi, J. Naughton, D. Roekaerts, B. Shader, S. Sritharan, and the Drs. H. Gopalan and M. Stöllinger. My sincere appreciation and thanks are expressed to all my colleagues at the Mathematics Department (University of Wyoming) for the pleasant atmosphere and a lot of help. I am very thankful to Professor G. Katul (Duke University, Durham, NC) for providing the instantaneous velocity and temperature data measured in the atmospheric surface layer (Chu et al. 1996), which were used for the illustration of real probability density functions in Sect. 4.5. Special thanks go to Dr. G. Turner (CSIRO, Sustainable Ecosystems, Australia) for providing the predictions of the World3 model, which describes the evolution of the global economic system (Turner 2008). These model predictions were used in Sect. 7.5 for the discussion of oscillations and collapse in population ecology. Many thanks also go to Dr. Ch. Baumann (Springer, Heidelberg) for his understanding and the good collaboration regarding the production of this text. For copyediting and careful text corrections I am grateful to Theodor C.H. Cole. And most of all, I want to deeply thank my wife Petra.

Stefan Heinz
Laramie, Wyoming
May 2011

# Contents

# 1 Deterministic Analysis of Observations

A usual modeling problem is given by the case that there are some relevant observations which may be obtained by measurements, and we would like to have a relatively simple mathematical function that provides a model for the observations. The purpose of finding such a function is to obtain a qualitative and quantitative understanding of observations, which is helpful for reacting to the observations in an appropriate way. The process of developing a mathematical model involves several basic steps. The first step is to present the problem as simple as possible, for example by transformations of data. The second step is to use modeling concepts to derive various reasonable models. The third step is to evaluate the models obtained in order to identify the optimal model. A model that is developed on the basis of observations represents the same information as given by the observations. However, a model should also provide an additional benefit. Thus, the fourth step is to demonstrate the advantage of the model development by deriving valuable conclusions that are not directly given by the observations.

The basic four steps of the modeling process will be illustrated in this chapter. We do only consider deterministic models here, and we only address the problem of developing models for observations that depend on one variable (models for several variables that can account for randomness will be considered in other chapters). The sort of problems considered in this chapter will be explained in Sect. 1.1. Sections 1.2–1.5 describe ways for dealing with the four basic steps of the mathematical modeling process. Section 1.2 explains data transformations for obtaining linear relations. Section 1.3 presents polynomial models that provide a basis for the development of simple analytical models. Section 1.4 addresses the evaluation of models by comparing four models for the development of the U.S. population. Section 1.5 addresses the modeling of global warming to show how models can be used for deriving conclusions regarding the driving mechanism for observations. The discussions in this chapter will be summarized in Sect. 1.6.

S. Heinz, *Mathematical Modeling*, DOI 10.1007/978-3-642-20311-4_1,

## 1.1 Motivation

**Modeling Problems.** Mathematical models can be most helpful for improving our understanding and dealing with relevant problems. Here are some examples:

- Human beings were always interested in motions of celestial objects such as stars, planets, comets, and galaxies – attempting to understand the reasons for the motion and paths of celestial objects. Such an understanding was helpful for agriculture, navigation, the making of calendars, and even astrology. Nowadays, astronomy (or astrophysics) is used to investigate the formation and development of the universe. Such studies may provide answers to important questions, e.g., regarding the existence of other intelligent life, the development of the solar system, and the ultimate fate of the universe.
- Energy supply is nowadays a prerequisite for maintaining the living conditions that we appreciate. Most people have made the stressful experience of a power outage: we are grateful when the light is back, refrigerator and computer working again. More importantly, available energy is the condition for many technical developments. The energy consumption does steadily grow, and we need quantitative knowledge of this development, this means a mathematical model for the energy consumption, to understand future needs.
- There are also many daily situations for which we will find it helpful to have a simple quantitative understanding of things. Suppose that you are driving your car. Given a certain velocity, what is an appropriate distance to the car in front of you such that you are able to stop safely? This distance will change with the car velocity, so it needs a simple formula for this calculation that you can easily use.
- Global warming is becoming a serious problem. An increase of the global temperature may cause glacial melting, Arctic shrinkage, and a worldwide sea level rise. Changes in the amount and pattern of precipitation may result in flooding and drought. There may be changes in the frequency and intensity of extreme weather events. Thus, an understanding of global warming is clearly relevant. In particular, we need mathematical models that explain human impacts on the global warming (e.g., as a consequence of greenhouse gas emissions).
- Other relevant mathematical modeling tasks arise from technical developments. The optimal design of technical processes often requires the simulation of complex processes that can hardly be studied otherwise. This concerns, e.g., the optimization of flow and chemical reactions in chemical reactors, and the flow around an aircraft or wind turbine. The ability to perform accurate oil reservoir simulations or $CO_2$ sequestration is a requirement, for example, to optimize the acquisition of natural resources.

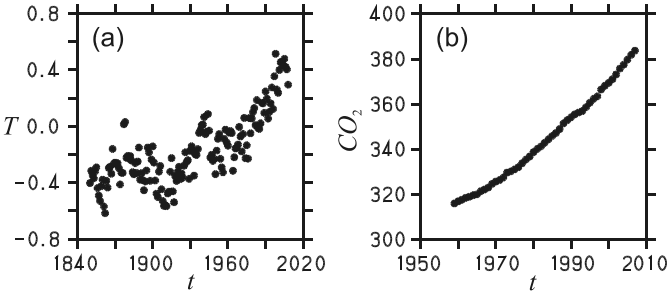



**Fig. 1.1.** An illustration of global warming. (a) The HadCRUT3 global temperature anomaly data (consisting of annual differences from 1961–90 normals) are given in ºC (Rayner et al. 2003, Brohan et al. 2006); (b) the Mauna Loa data (Tans 2008) of atmospheric $CO_2$ concentrations are given in ppmv.

- Another example that is clearly relevant to daily life is given by the weather prediction. It is helpful to know the probability for significant temperature changes, rain or snow, and we need knowledge about the developments of blizzards, hurricanes, and tornadoes to be prepared for their consequences. Such weather predictions have to be performed by means of numerical simulations, and the basis for such simulations is given by mathematical models.

**Problems Considered.** The question of how it is possible to develop models for the first four problems described in the preceding paragraph will be addressed in the following sections of this chapter. The last two problems require the use of very advanced techniques that cannot be presented in this book. A typical problem considered here is illustrated in Fig. 1.1 that illustrates the global warming in terms of the increase of the atmospheric $CO_2$ concentrations and the related increase of the global temperature anomaly $T$. A quantitative understanding of this relevant problem, including the explanation of the reason for the observed global warming, requires the development of mathematical models for these data trends and the analysis of the relation of these trends. A solution to this question will be presented in Sect. 1.5.

## 1.2 Data Transformations: Linear Models

First, let us devote ourselves to the simplest modeling approach: the application of linear functions. It is often the case that data considered do not follow a linear function. However, data transformations sometimes allow the use of linear models for redefined data. Several examples for this approach will be given in the following. Other examples will be considered in Sects. 1.4 and 1.5.

| $t$ | $C$ | $t$ | $C$ |
|---|---|---|---|
| 1950 | 34.616 | 1980 | 78.122 |
| 1955 | 40.208 | 1985 | 76.491 |
| 1960 | 45.087 | 1990 | 84.652 |
| 1965 | 54.017 | 1995 | 91.173 |
| 1970 | 67.844 | 2000 | 98.975 |
| 1975 | 71.999 | 2005 | 100.506 |

**Table 1.1** The U.S. energy consumption $C$ (in $10^{15}$ Btu) in time $t$ (U.S. Dept. of Energy 2008).

### *1.2.1 Energy Consumption*

The U.S. Department of Energy (2008) published data on the U.S. energy consumption in its Annual Energy Report 2007. The data are given from 1950 to 2005 in Table 1.1 and shown in Fig. 1.2. Let us develop a model for these data.

**Linear Energy Consumption Model.** The data presented in Fig. 1.2 support the use of a linear function for the modeling of the energy consumption. A linear model for the energy consumption $C$ that passes any given data points $(t_1, C_1)$ and $(t_2, C_2)$ can be written as

$$C = C_1 \frac{t - t_2}{t_1 - t_2} + C_2 \frac{t - t_1}{t_2 - t_1}. \tag{1.1}$$

The validity of this expression can be proven by considering $C$ at $t_1$ and $t_2$. This shows that $C(t_1) = C_1$ and $C(t_2) = C_2$, as required. To represent the overall data trend, it appears to be reasonable to use a linear function that passes the points (1950, 34.6) and (1995, 91). The use of these data in Eq. (1.1) leads to

$$C = 34.6 \frac{t - 1995}{1950 - 1995} + 91 \frac{t - 1950}{1995 - 1950}. \tag{1.2}$$

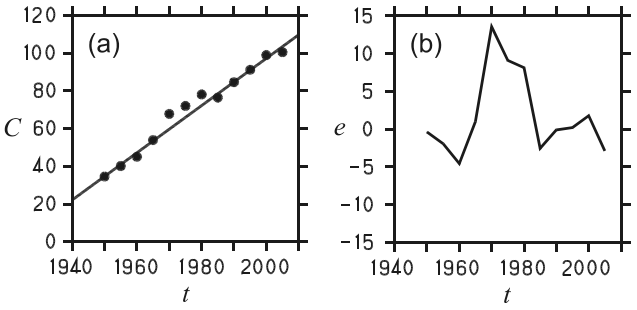



**Fig. 1.2.** U.S. energy consumption $C$ (in $10^{15}$ Btu) in time $t$. (a) The data from Table 1.1 are shown as *dots*, and the *line* represents the linear model (1.3); (b) shows the relative error $e$ (in %) of the model (Eq. 1.3).

A convenient way to write this formula is to reformulate both contributions by referring to $t - 1995$. Thus, Eq. (1.2) becomes

$$C = 91 + 1.25(t - 1995). \tag{1.3}$$

This writing of the model for the energy consumption reveals the value $C = 91$ at the time $t = 1995$ and the rate of the consumption increase (the slope 1.25).

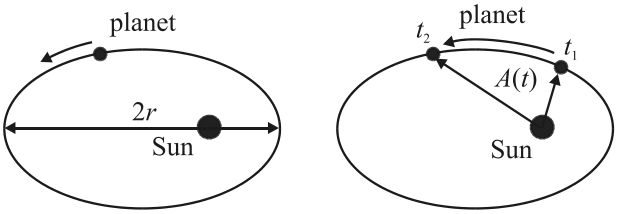
**Model Evaluation.** Figure 1.2 shows the good performance of this model. The trend of the energy consumption is well represented, and the model function agrees relatively well with the data points. The model performance can be seen in more detail by calculating the relative error. Given any set of data points $(x_1, y_1)$, $(x_2, y_2)$, $\cdots$ $(x_n, y_n)$, the error $e$ related to each data point can be defined by

$$e = \frac{y - y^{(\mathrm{mod})}}{y^{(\mathrm{mod})}}. \tag{1.4}$$

Here, $y^{(\mathrm{mod})}$ refers to the value predicted by the model. This value is considered to be the true value. Obviously, this view depends on the performance of the model, this means a model that clearly disagrees with any observed data trend cannot be considered to provide a true value. The symbol $y$ denotes the observed data point. This value is considered to be affected by randomness. The relative error related to the use of the model (1.3), $y = C$ in our case, is shown in Fig. 1.2b. This figure shows that the maximum relative error of Eq. (1.3) is given by 13.5% at $t = 1970$. Such an error means that Eq. (1.3) represents a reasonable model for the energy consumption for the period considered. It is worth noting that the magnitude of the relative error is smaller than 3% for $t > 1985$. Such error values are typical for a good model. An attractive feature of the model given by Eq. (1.3) is its simplicity, which enables an easy understanding: the energy consumption increases linearly in time $t$ with an increase rate of 1.25. The model is certainly useful. It can be used, for example, for making reasonable predictions of the energy consumption for the next decade.

### *1.2.2 Kepler's Third Law*

Let us consider another example. The German mathematician and astronomer Johannes Kepler (1571–1630) analyzed over 20 years the motion of planets on the basis of the astronomical observations of the Danish astronomer Tycho Brahe. By 1609 Kepler had formulated his first two laws of planetary motion: (K1) a planet revolves around the Sun in an elliptical orbit with the Sun at one focus, and (K2) the line joining the Sun to a planet sweeps out equal areas $A(t)$ in equal times (i.e., $dA/dt$ is constant, see the illustration in Fig. 1.3). Kepler then spent many years to




(a) Kepler's First Law (b) Kepler's Second Law

**Fig. 1.3.** An illustration of Kepler's First Law and Second Law.

address the question of what determines the orbital period $T_P$ (the number of Earth years to orbit the Sun with respect to background stars) of the revolution of a planet. The latter question is relevant to see, for example, whether there are laws that govern the planetary motion.

**Linearized Model.** Let us use the data in Table 1.2 to address the same question: how does $T_P$ depend on the mean distance $r$ from the Sun (the mean distance is the sum of the maximal and minimal distance divided by two)? Figure 1.4a shows that a linear model clearly disagrees with the data trend: the data increase faster than a linear function. Figure 1.4b shows that a quadratic function (which can be found by adopting the approach presented in Sect. 1.3) also disagrees with the data trend: this function increases faster than the data. The behavior of linear and quadratic models indicates that $T_P$ can be a power function of r given by

$$T_P = A r^B. \tag{1.5}$$

Here, $A$ and $B$ are any constants. The best way to test the validity of this assumption is to consider a corresponding linear relation between $\ln T_P$ and $\ln r$, which follows from taking the natural logarithm of Eq. (1.5),

$$\ln T_P = \ln A + B \ln r. \tag{1.6}$$

To prove the validity of this assumption we have to plot $\ln T_P$ against $\ln r$. Indeed, Fig. 1.5a provides evidence for the suitability of such a linear relation. By using the two-point formula (1.1) we find (the second and the seventh point are used)

$$\ln T_P = 2.8490 + 1.4996 \ln r. \tag{1.7}$$

| Planet | $r$ | $T_P$ |
|---|---|---|
| Mercury | 0.0579 | 0.2408 |
| Venus | 0.1082 | 0.6152 |
| Earth | 0.1496 | 1.0000 |
| Mars | 0.2280 | 1.8808 |
| Jupiter | 0.7785 | 11.8618 |
| Saturn | 1.4335 | 29.4566 |
| Uranus | 2.8718 | 84.0107 |
| Neptune | 4.4948 | 164.7858 |

**Table 1.2** Orbital periods and mean distances of planets from the Sun (World Almanac 2010). Here, $r$ is the mean distance from the Sun (which is equal to the major axis of the ellipse) in $10^9$ km. $T_P$ is the period in earth years a = 365.256 days, this means $T_P$ gives the number of Earth years to orbit the Sun with respect to background stars.

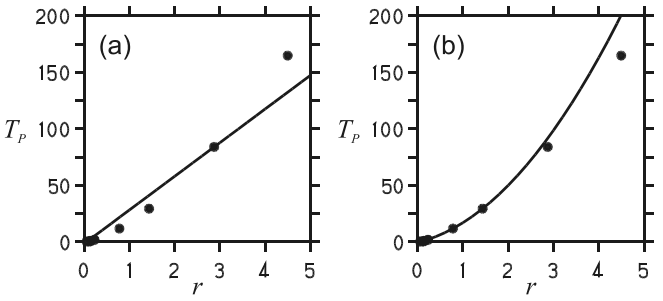



**Fig. 1.4.** The orbital period $T_P$ of planets in dependence on the mean distance $r$ from the Sun. The data from Table 1.2 (World Almanac 2010) are given by *dots*. The *lines* in (a) and (b) show linear and quadratic functions for $T_P$, respectively.

By taking both sides of this relation as exponents of an exponential function, the dependence of $T_P$ on $r$ can be written as

$$T_P = \sqrt{\frac{4\pi^2}{G_S} r^3}. \tag{1.8}$$

This is Kepler's Third Law (K3): the square of the period of revolution of a planet is proportional to the cube of the major axis of its orbit. Here, $G_S = 0.1324$ $(10^9 \text{ km})^3 / \text{a}^2$ is the standard gravitational parameter. The constant $G_S$ can also be written as $G_S = 1.3291\ 10^{20}\ \text{m}^3/\text{s}^2$, where 1 km = $10^3$ m and a = 31,558,118.4 s.

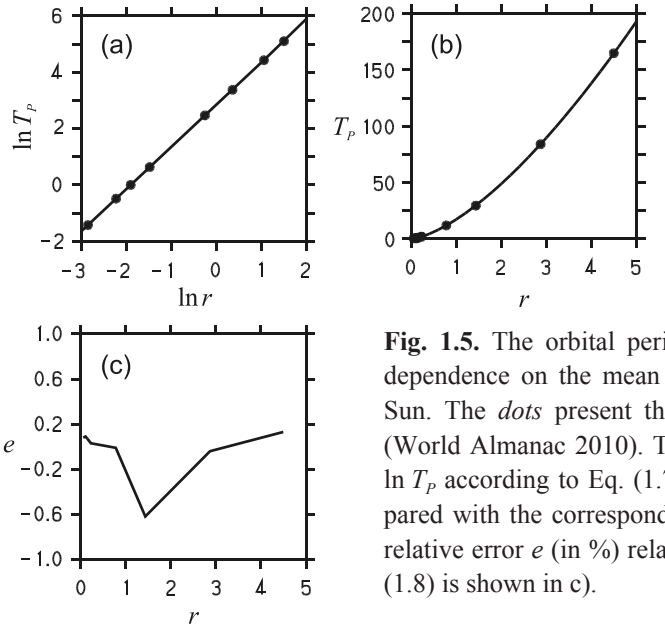



**Fig. 1.5.** The orbital period $T_P$ of planets in dependence on the mean distance $r$ from the Sun. The *dots* present the data of Table 1.2 (World Almanac 2010). The *line* in (a) shows $\ln T_P$ according to Eq. (1.7). Eq. (1.8) is compared with the corresponding data in (b). The relative error $e$ (in %) related to the use of Eq. (1.8) is shown in c).

| $v$ | $D$ | $v$ | $D$ |
|---|---|---|---|
| 10 | 27 | 60 | 303 |
| 15 | 44 | 65 | 344 |
| 20 | 63 | 70 | 387 |
| 25 | 85 | 75 | 433 |
| 30 | 109 | 80 | 481 |
| 35 | 135 | 85 | 531 |
| 40 | 164 | 90 | 584 |
| 45 | 195 | 95 | 639 |
| 50 | 229 | 100 | 696 |
| 55 | 265 | | |

**Table 1.3** The characteristic total stopping distance $D$ (in feet) related to the automobile's velocity $v$ (in mph) according to the Code of Virginia (Jernigan & Kodaman 2001).

**Model Evaluation.** Figure 1.5b shows that Eq. (1.8) agrees very well with the data: there is no observable difference between them. This fact indicates that Eq. (1.8) represents a very accurate model. Evidence for this view is provided by Fig. 1.5c that shows the relative error $e$ in % related to the use of Eq. (1.8). As may be seen, the relative error is very small (the magnitude is smaller than 0.6%). It should be noted that there is no unique way to calculate the parameter $G_S$ and the power 1.5 of $r$ in Eq. (1.8). Very little modifications of these values, e.g., a power 1.4996, would imply a model with the same accuracy. However, the power will be always found very close to 1.5. Thus, the value 1.5 is the most appropriate value because it indicates a deep theoretical relationship between $T_P$ and $r$. The value $G_S = 1.3291 \times 10^{20}$ m$^3$ / s$^2$ provides the best agreement between the model and data. The simplicity of Eq. (1.8) is very helpful for the understanding. The observation that the orbital period $T_P$ is fully controlled by the mean distance $r$ from the Sun means that $T_P$ is independent of properties (e.g., of the mass) of planets. This feature indicates the possible existence of a universal law that governs the planetary motion. Indeed, it will be shown in Chap. 3 that Newton's Law of Gravitation can be derived on the basis of Kepler's Third Law.

### *1.2.3 Vehicular Stopping Distance*

Let us consider a problem from daily life: the vehicular stopping distance. Say you drive your car at a certain velocity $v$, and then, you need to come to a full stop. The stopping distance $D$ is the distance that you will still drive after trying to stop. Table 1.3 shows data for the dependence of $D$ on $v$ according to the current Code of Virginia (Jernigan & Kodaman 2001). Knowledge about the stopping distance $D$ is relevant to keep a good distance to the car in front of you. How is it possible to model the dependence of $D$ on $v$? In particular, how can we derive a formula for $D(v)$ that we can easily use?

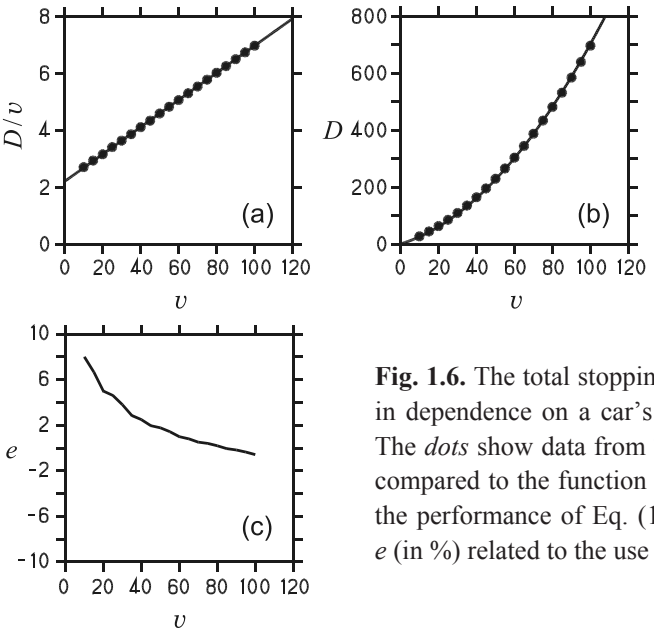



**Fig. 1.6.** The total stopping distance $D$ (in feet) in dependence on a car's velocity $v$ (in mph). The *dots* show data from Table 1.3. (a) $D/v$ is compared to the function $2.2 + v/21$ (*line*); (b) the performance of Eq. (1.9); (c) relative error $e$ (in %) related to the use of Eq. (1.9).

**Linearized Model.** A reasonable approach is to think about the suitability of a linear function. However, Fig. 1.6 shows that a linear function cannot represent an accurate model. A way to see the deviations from a linear function is to plot $D/v$ against $v$. This approach involves the assumption that $D$ is proportional to $v$. This view is correct because $D$ should be equal to zero for $v = 0$. A plot of $D/v$ is shown in Fig. 1.6a. Obviously, a linear function for $D/v$ represents a very good model. Instead of using $D = (2.2 + v/21)\,v$, which follows from the two-point formula (1.1) combined with the points (15, 44/15) and (95, 639/95), the function

$$D = \left(2 + \frac{v}{20}\right) v \tag{1.9}$$

will be applied here. The main reason for doing this is that we are interested in a model that can be easily applied for the calculation of $D$.

**Model Evaluation.** Figure 1.6b shows that Eq. (1.9) represents a very good model. A closer look reveals the following: The trend of the data is well represented, and there is a good agreement between the model and data points. The accuracy of Eq. (1.9) can be evaluated by considering its relative error $e$ shown in Fig. 1.6c. We see that the error magnitude is smaller than 1% for $v > 60$ mph. For smaller velocities there are error magnitudes of up to 8% (at $v = 10$ mph). However, it is more important to assess the stopping distance at higher velocities. Thus, the model accuracy is excellent. The model is simple, which is helpful for using it for the calculation of stopping distances including velocities which are

higher than the range of $v$ considered in Fig. 1.6. Is the model understandable? We may expect that $D$ depends, basically, linearly on $v$. Why do we observe a quadratic contribution $v^2$ in Eq. (1.9)? The stopping distance $D$ must depend on the kinetic energy $m\,v^2/2$ of the car ($m$ is the car's mass): the higher the kinetic energy the longer it will take it to stop. Therefore, a quadratic dependence of $D$ on $v$ is reasonable.

## 1.3 Model Development: Polynomial Models

The approach presented in Sect. 1.2 is extremely helpful because of its simplicity, but it has a rather limited range of applicability: most models cannot be developed in terms of linear functions. To deal with this problem we will consider polynomial models. Their use represents a flexible strategy for the development of relatively simple models that can represent a variety of trends.

### *1.3.1 The Lagrangian Form of Polynomials*

**Linear Polynomials.** First, we consider linear polynomials, i.e., a function $y = a_0 + a_1\,x$. Here, $x$ is the independent variable, and $y$ is the dependent variable (any variable for which we need a model). The function $y = a_0 + a_1\,x$ involves two unknown parameters: $a_0$ and $a_1$. The parameters can be calculated by the constraint that $y(x)$ passes through two points $(x_1, y_1)$ and $(x_2, y_2)$, which we consider to be given by observations. The requirement that $y(x)$ passes through these two points implies the following two relations,

$$y_1 = a_0 + a_1 x_1, \qquad y_2 = a_0 + a_1 x_2. \tag{1.10}$$

These relations can be used to calculate the parameters $a_0$ and $a_1$ in $y = a_0 + a_1\,x$. This calculation provides $y = P_1(x)$, where

$$P_1(x) = y_1 \frac{x - x_2}{x_1 - x_2} + y_2 \frac{x - x_1}{x_2 - x_1}. \tag{1.11}$$

Here, $P_1(x)$ refers to a polynomial of first order. The term polynomial refers to a function that consists of the sum or difference of integer powers of $x$ multiplied with any constants. The subscript one of $P_1(x)$ refers to a polynomial of first order. It is assumed that the positions $x_1$ and $x_2$ are different to avoid division by zero. How is it possible to prove the claim that $y = P_1(x)$ passes through the two points considered? The simplest way is to consider $P_1(x)$ at $x_1$ and $x_2$. In this way, one finds the identities $y(x_1) = y_1$ and $y(x_2) = y_2$, as required.

**Quadratic Polynomials.** Second, we consider quadratic polynomials, this means a function $y = a_0 + a_1 x + a_2 x^2$. This function $y(x)$ is completely determined by the constraint that $y(x)$ passes through three given points $(x_1, y_1)$, $(x_2, y_2)$ and $(x_3, y_3)$. The latter constraint provides the relations

$$y_1 = a_0 + a_1 x_1 + a_2 x_1^2, \quad y_2 = a_0 + a_1 x_2 + a_2 x_2^2, \quad y_3 = a_0 + a_1 x_3 + a_2 x_3^2 \tag{1.12}$$

for the calculation of the parameters $a_0$, $a_1$, and $a_2$. The function $y(x)$ obtained in consistency with these constraints can be written $y = P_2(x)$, where

$$P_2(x) = y_1 \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)} + y_2 \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)} + y_3 \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)}. \tag{1.13}$$

The positions $x_1$, $x_2$, and $x_3$ are assumed to be different. The validity of this expression can be proven by considering $P_2(x)$ at $x_1$, $x_2$, and $x_3$, which provides the identities $y(x_1) = y_1$, $y(x_2) = y_2$, and $y(x_3) = y_3$.

**Lagrangian Form of Polynomials.** Third, we consider the general case that we have any polynomial of order $n$. To define such a polynomial uniquely, we assume that observations are given at $n + 1$ points, this means $(x_0, y_0)$, $(x_1, y_1)$, $\cdots$ $(x_n, y_n)$. The reason for using $n + 1$ points for the definition of a polynomial of order $n$ can be seen by considering again linear and quadratic polynomials: we need two points to define a linear function, and we need three points to define a quadratic function. The way to construct a polynomial $y = P_n(x)$ becomes clear by having a closer look at the structure of linear and quadratic functions. Let us consider Eq. (1.13) for a quadratic polynomial. For every $y_i$ ($i = 1, 3$), the numerators involve products of differences between $x$ and all the other positions except the difference $x - x_i$. The denominators are equal to the numerators except that $x$ is replaced by $x_i$. Hence, a polynomial $y = P_n(x)$ of order $n$ is given by

$$P_n(x) = y_0 L_0(x) + y_1 L_1(x) + \cdots y_n L_n(x), \tag{1.14}$$

where the functions $L_k(x)$ are defined by ($k = 0, n$)

$$L_k(x) = \prod_{i=0,\, i \neq k}^{n} \frac{x - x_i}{x_k - x_i} = \frac{(x - x_0)(x - x_1) \cdots (x - x_{k-1})(x - x_{k+1}) \cdots (x - x_n)}{(x_k - x_0)(x_k - x_1) \cdots (x_k - x_{k-1})(x_k - x_{k+1}) \cdots (x_k - x_n)}. \tag{1.15}$$

Here, it is assumed that the positions $x_1, x_2, \cdots, x_n$ are different to avoid division by zero. The correctness of this formula can be seen by considering $P_n(x)$ at $x_0, x_1, \cdots, x_n$. This shows that the polynomial $y = P_n(x)$ passes through all the observations $(x_0, y_0)$, $(x_1, y_1)$, $\cdots$ $(x_n, y_n)$. The polynomial formula (1.14) is called the Lagrangian form of polynomials. The settings $n = 1$ and $n = 2$ reveal that the polynomial (1.14) generalizes the expressions for linear and quadratic functions, respectively.

**Table 1.4** Data sets used for the illustration of properties of exact polynomials.

| $x$ | (*a*) $y$ | (*b*) $y$ | (*c*) $y$ | (*d*) $y$ | (*e*) $y$ | (*f*) $y$ |
|---|---|---|---|---|---|---|
| 1 | 1.0001 | 1.001 | 1.01 | 1.05 | 1.00 | 1.00 |
| 2 | 1.9998 | 1.998 | 1.98 | 1.90 | 2.00 | 2.00 |
| 3 | 3.0003 | 3.003 | 3.03 | 3.15 | 3.00 | 3.00 |
| 4 | 3.9996 | 3.996 | 3.96 | 3.80 | 4.04 | 4.80 |
| 5 | 5.0005 | 5.005 | 5.05 | 5.25 | 5.00 | 5.00 |
| 6 | 5.9994 | 5.994 | 5.94 | 5.70 | 6.00 | 6.00 |
| 7 | 7.0007 | 7.007 | 7.07 | 7.35 | 7.00 | 7.00 |
| 8 | 7.9992 | 7.992 | 7.92 | 7.60 | 8.00 | 8.00 |

### *1.3.2 Properties of Polynomials*

**Exact Polynomial Models.** Polynomial models have many advantages. Polynomials can be easily differentiated and integrated. Analytical derivatives and integrals that are obtained in this way are helpful to find, for example, minimum and maximum values of variables or integral values like work and probability. For a given number of observations it is always possible to construct a polynomial that passes exactly through all the data points ($n + 1$ data points define a polynomial of order $n$). Such a polynomial will be called here an exact polynomial model. However, it turns out that models obtained in this way are often not very useful, the reason being that observations are always affected by errors, and such errors may imply a poor performance of exact polynomial models. Let us have a closer look at this problem.

**Erroneous Data.** Data used for the development of models may be affected by a variety of errors. It is often the case that the conditions for experiments are affected by little changes. For temperature measurements in the atmosphere, for example, one does never find exactly the same conditions. Such little changes of conditions for experiments will produce noisy observations. The way in which data are recorded does also induce errors. Every measurement method has a certain error: there is usually no way to perform absolutely correct measurements. With regard to measurements of human populations, it is often only possible to obtain good estimates for the real numbers (e.g., for the number of residents of countries or cities). Another source of errors is given by the fact that observations (e.g., temperature measurements) have to be presented numerically with a certain number of digits. Round-off errors that are introduced in this way will also add errors to data sets. Occasionally, single measurements are simply incorrect because of a variety of possible recording problems.

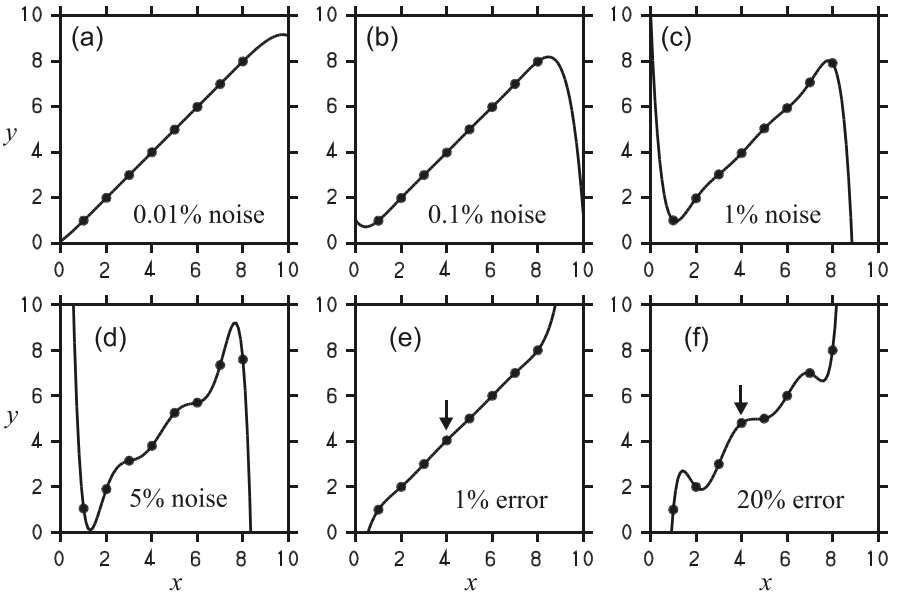



**Fig. 1.7.** Exact polynomial models of 7th order for the six cases of data points given in Table 1.4. The data points are given by *dots*. The polynomial models are given by *lines*. All the curves pass exactly through the data points. The *arrows* in (e) and (f) indicate the position of incorrect values.

**Effect of Erroneous Data.** Let us consider the data presented in Table 1.4 to illustrate the effect of erroneous data on exact polynomial models. The cases (a) and (b) present noisy $y$ data with a noise of 0.01% and 0.1%, respectively. These cases illustrate the influence of round-off errors. The cases (c) and (d) present noisy $y$ data with a noise of 1% and 5%, respectively. Such data may result from changing conditions for experiments or the inaccuracy of measurement methods. The cases (e) and (f) involve incorrect $y$ values at $x = 4$ (errors of 1% and 20%, respectively). The implications of these data sets are illustrated in Fig. 1.7 that shows exact polynomial models of 7th order for these six cases. These polynomials can be obtained by specifying the corresponding Lagrangian form of polynomials (1.14). Figures 1.7a and 1.7b reveal that even very minor noise effects of 0.01% and 0.1% may have significant effects on exact polynomial models: such polynomials are very sensitive to small changes in the data. The trend of data is given by a linear function $y = x$. Instead, the polynomials show severe oscillations after the interval for which data are given. Obviously, the use of such models to extrapolate data trends may be completely wrong. Figures 1.7c and 1.7d show that

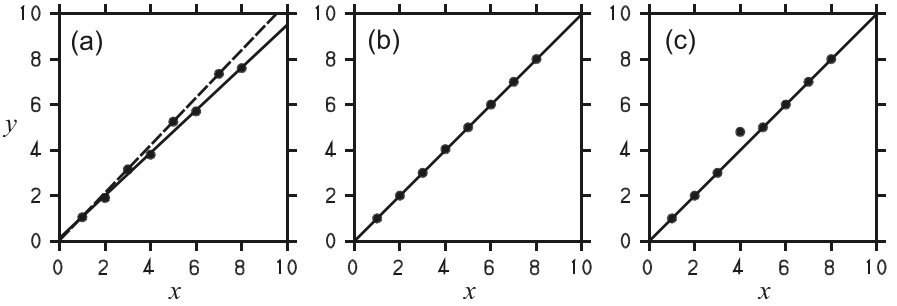



**Fig. 1.8.** Reduced polynomial models of first order for the cases (d), (e), and (f) of Table 1.4. The data are given by *dots*, and polynomials are given by *solid lines*. The polynomials are defined by the constraint to pass through the first and last data point. The *dashed line* in (a) shows a polynomial that passes through the first point and the point next to the last point.

these effects are more pronounced for higher noise intensities. In addition to the curve behavior in Figs. 1.7a–b, the model behavior between the data points is now also incorrect: see the curve for $1 \le x \le 2$ and $7 \le x \le 8$ in Fig. 1.7d. The curves present all the data points correctly, but there is no guarantee that the model is correct for any deviations from the data points. Figures 1.7e and 1.7f show that even relatively small deviations from correct recordings of data points may have severe implications for the performance of models. One erroneous recording with an error of 1% means that the model becomes unusable for the extrapolation of data. One erroneous recording with an error of 20% results in a model that is not useful at all.

**Reduced Polynomial Models.** A way to avoid the extreme sensitivity of exact polynomials to small changes in the data is to apply low-order polynomials (e.g., linear, quadratic, or cubic polynomials). Such polynomials are constructed on the basis of a reduced number of data points. The models obtained in this way are called here reduced polynomial models. The properties of reduced polynomials are illustrated in Fig. 1.8. These functions are capable of representing the correct data trend. Low-order polynomials are more stable than exact polynomials: their use is related to a smoothing. Reduced polynomial models have, however, a major disadvantage because there is no unique way to construct such models. These models depend on the data chosen to define the polynomial considered. Figure 1.8a illustrates that the use of different data points used to define such polynomials may have a significant effect on the model.

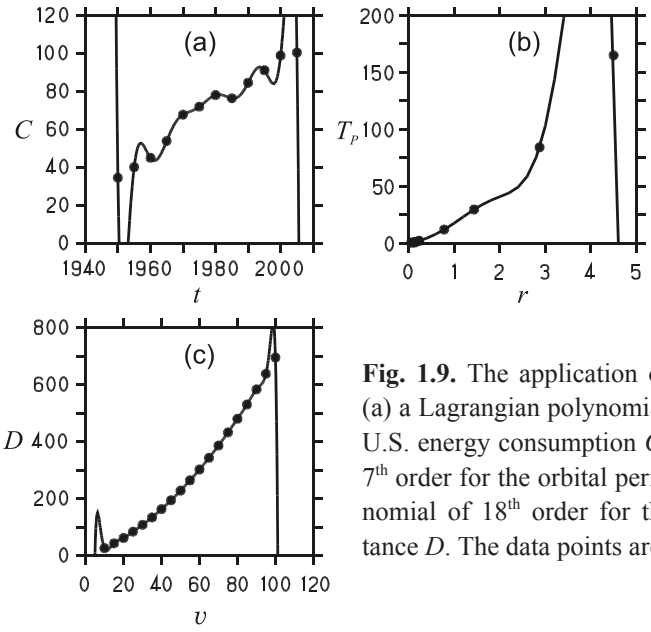



**Fig. 1.9.** The application of exact polynomials: (a) a Lagrangian polynomial of 11th order for the U.S. energy consumption $C$, (b) a polynomial of 7th order for the orbital period $T_P$, and (c) a polynomial of 18th order for the total stopping distance $D$. The data points are given by *dots*.

### *1.3.3 Polynomial Models of Observations*

Let us reconsider the examples discussed in Sect. 1.2 to illustrate the performance of polynomials regarding the modeling of real data.

**Exact Polynomial Models.** Figure 1.9a presents a Lagrangian polynomial of 11th order as a model for the U.S. energy consumption $C$. This model provides an exact curve through the 12 data points considered (the data between 1950 and 2005). The model shows severe oscillations between the data points. The predicted trend after 2005 is clearly in disagreement with the data trend. Thus, this approach does not work at all. Figure 1.9b shows the disadvantages of exact polynomials for the modeling of the orbital period $T_P$. The polynomial of 7th order passes all the data points, but its behavior is incorrect for $r \geq 2.9$. Figure 1.9c shows an exact polynomial of 18th order as a model for the stopping distance $D$. This polynomial provides an accurate model except for the behavior of $D$ close to the end points of the interval considered (see the curve between $v = 95$ mph and $v = 100$ mph). The oscillations observed indicate again the disadvantages related to the use of an exact polynomial model. Such a model is useless for the prediction of $D$ at velocities $v > 100$ mph. Apart from that, knowledge of a polynomial of 18th order will not help a lot to find an appropriate distance to the car in front of you: the use of such a polynomial is not easy!

**Table 1.5.** The development of the U.S. population in time from 1790–2050 according to the Decennial Censuses, U.S. Census Bureau, U.S. Dept. of Commerce (World Almanac 2010). The population $P$ is measured in $10^9$ and $t$ refers to the year. The values for $t > 2000$ are projections (Population Division, U.S. Census Bureau, NP2008-T1, August 14, 2008).

| $t$ | $P$ | $t$ | $P$ | $t$ | $P$ |
|---|---|---|---|---|---|
| 1790 | 0.0039 | 1880 | 0.0502 | 1970 | 0.2033 |
| 1800 | 0.0053 | 1890 | 0.0630 | 1980 | 0.2265 |
| 1810 | 0.0072 | 1900 | 0.0762 | 1990 | 0.2487 |
| 1820 | 0.0096 | 1910 | 0.0922 | 2000 | 0.2814 |
| 1830 | 0.0129 | 1920 | 0.1060 | 2010 | 0.3102 |
| 1840 | 0.0171 | 1930 | 0.1232 | 2020 | 0.3414 |
| 1850 | 0.0232 | 1940 | 0.1321 | 2030 | 0.3735 |
| 1860 | 0.0314 | 1950 | 0.1513 | 2040 | 0.4057 |
| 1870 | 0.0398 | 1960 | 0.1793 | 2050 | 0.4390 |

**Reduced Polynomial Models.** The use of reduced polynomial models for the modeling of the U.S. energy consumption $C$, the orbital period $T_P$, and the total stopping distance $D$ shows different features. After rewriting the orbital period and stopping distance data, it was shown in Sect. 1.2 that all the three problems can be successfully solved on the basis of linear functions. Further examples for the usefulness of reduced polynomial models will be discussed in Sects. 1.4 and 1.5. Thus, the application of reduced polynomial models is in general much more helpful than the use of exact polynomial models.

## 1.4 Model Evaluation: Population Modeling

Two approaches for creating a model were considered so far: the development of linear models (for transformed data if required) in Sect. 1.2 and the use of polynomial models in Sect. 1.3. Next, let us consider the relevant question of how a model can be evaluated. This question will be addressed by considering the development of the U.S. population in time according to the Decennial Censuses, U.S. Census Bureau, U.S. Dept. of Commerce (World Almanac 2010). The data are given in Table 1.5. First, we will address the modeling of the U.S. population in terms of linear models for redefined data. This approach is driven by concepts for the modeling of population dynamics that will be described in detail in Sect. 7.4 (here we use these concepts simply as functions that can be transformed to linear models for redefined variables). Second, we will consider the application of polynomial models. The second approach is driven by data – we try to find an appropriate model as an interpolation and extrapolation of available data. Third, we evaluate the suitability of these two modeling approaches for the development of the U.S. population.

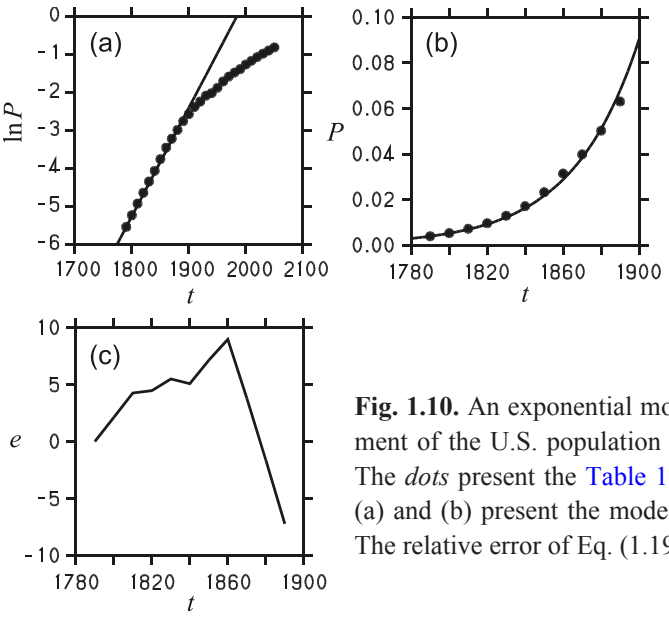



**Fig. 1.10.** An exponential model for the development of the U.S. population $P$ (in $10^9$) in time $t$. The *dots* present the Table 1.5 data. The *lines* in (a) and (b) present the models (1.18) and (1.19). The relative error of Eq. (1.19) is shown in c).

### *1.4.1 Linearized Conceptual Models*

Several concepts for the modeling of population dynamics will be discussed in Sects. 7.4 and 7.5. Here, we will focus on the application of two models that can be written as linear models for redefined variables: the exponential model (7.82), which is called the Malthusian Law, and the logistic model (7.102).

**Exponential Growth.** The Malthusian Law (7.82) represents the assumption that populations grow exponentially:

$$P = P_0\, e^{r(t-t_0)}. \tag{1.16}$$

Here, $P_0$ is the initial value of $P$ at $t_0$, and $r$ is the growth rate. Equation (7.82) is used here in a slightly modified manner because of the consideration of a nonzero initial time $t_0$. The most convenient way to look at the Malthusian Law is to write it as a linear relation between $\ln P$ and $t$,

$$\ln P = \ln P_0 + r(t - t_0). \tag{1.17}$$

The natural logarithm of the population data is shown as a function of $t$ in Fig. 1.10a to see the suitability of this assumption. Hence, $\ln P$ can be well described by a linear function of $t$ for the population data from 1790 to 1890. The parameters of the linear function are found by using the two-point formula (1.11).

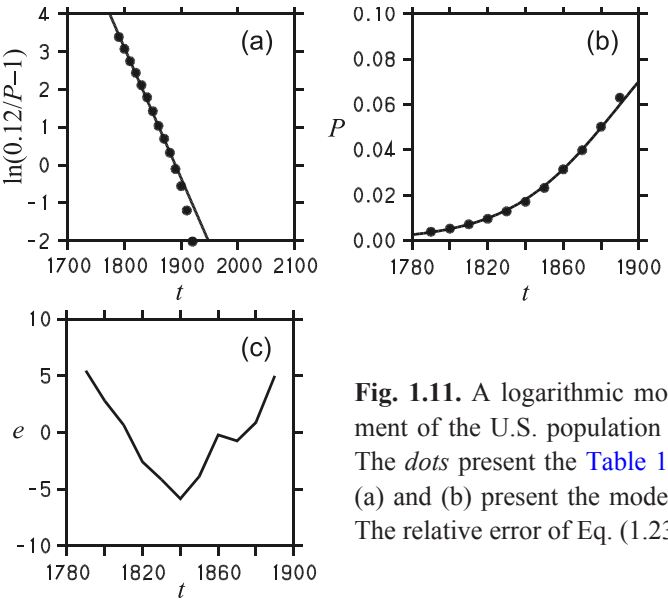



**Fig. 1.11.** A logarithmic model for the development of the U.S. population $P$ (in $10^9$) in time $t$. The *dots* present the Table 1.5 data. The *lines* in (a) and (b) present the models (1.22) and (1.23). The relative error of Eq. (1.23) is shown in c).

The linear function obtained in this way is given by

$$\ln P = \ln 0.0039 + \frac{t-1790}{35}. \tag{1.18}$$

Here, the initial value $P(1790) = 0.0039$ is used according to the data in Table 1.5, and the growth rate $r = 1/35$ provides the correct increase of population data. The latter model implies the exponential model

$$P = 0.0039\, e^{(t-1790)/35}. \tag{1.19}$$

Figure 1.10b shows a good agreement with the model and data. Fig. 1.10c shows that the magnitude of the relative model error is smaller than 9%.

**Logistic Growth.** Another modeling concept is to assume a logistic growth. The logistic model can be written in several ways (see Sect. 7.4.2). We will use the centered logistic model of Eq. (7.104),

$$P = P_c\left(1 + \frac{1 - e^{-(t-t_c)/\tau}}{1 + e^{-(t-t_c)/\tau}}\right) = \frac{2\,P_c}{1 + e^{-(t-t_c)/\tau}}. \tag{1.20}$$

Here, $P_c = P(t_c)$, $t_c$ is the critical point at which $dP/dt$ has a maximum, and $\tau$ refers to a characteristic time for the transition to the equilibrium state. In particular, we use Eq. (7.104) here in a simplified version by neglecting a constant $P_{-\infty}$ that may be added to $P$ (there is no need to involve such an additional constant $P_{-\infty}$ here). The best way to prove the suitability of the population model (1.20) is again to look at the linear relationship between $\ln(2P_c / P - 1)$ and $t$ that is implied by Eq. (1.20),

$$\ln\left(2P_c / P - 1\right) = -\frac{t - t_c}{\tau}. \tag{1.21}$$

To compare the logistic model with the exponential model we consider again the population data from 1790 to 1890. Figure 1.11a shows that $\ln(0.12 / P - 1)$ can be modeled by a linear function:

$$\ln\left(0.12 / P - 1\right) = -\frac{t - 1890}{29}. \tag{1.22}$$

The rewriting of this relation results in the logarithmic population growth model

$$P = \frac{0.12}{1 + e^{-(t-1890)/29}}. \tag{1.23}$$

The parameters of this model are chosen such that $P(1890) = 0.06$ according to the data in Table 1.5, and the growth rate 1/29 provides the correct increase rate. Figure 1.11b shows good agreement between the model and the data. This view is supported by Fig. 1.11c that reveals the magnitude of the relative model error to be smaller than 5%.

### *1.4.2 Polynomial Models*

An alternative approach for the model development is the use of polynomials discussed in Sect. 1.3. In agreement with the discussion of the properties of exact polynomial models, it turns out that the use of a polynomial of 26th order that provides an exact curve through the 27 data points given in Table 1.5 is pointless: a reasonable model cannot be obtained in this way. Thus, we have to use reduced polynomial models. Here, we will use linear and quadratic models.

**Quadratic Polynomial.** First, we will consider the suitability of a quadratic polynomial. According to Eq. (1.13), the quadratic polynomial is given by

$$P = P_1 \frac{(t - t_2)(t - t_3)}{(t_1 - t_2)(t_1 - t_3)} + P_2 \frac{(t - t_1)(t - t_3)}{(t_2 - t_1)(t_2 - t_3)} + P_3 \frac{(t - t_1)(t - t_2)}{(t_3 - t_1)(t_3 - t_2)}. \tag{1.24}$$

The population data at 1790, 1890, and 2040 are taken as the three data points required for the use of this formula. The performance of the quadratic polynomial is illustrated in Fig. 1.12a. This polynomial describes the population data over the entire data range considered. The model error is given in Fig. 1.12b. The error values have a peak of –32% at $t = 1810$. However, the population values are very small in this range so that such errors are acceptable. For $t \geq 1840$, the error values are found to be relatively small, $-3.4\% < e < 8.5\%$. Thus, the second-order polynomial model represents a good model.

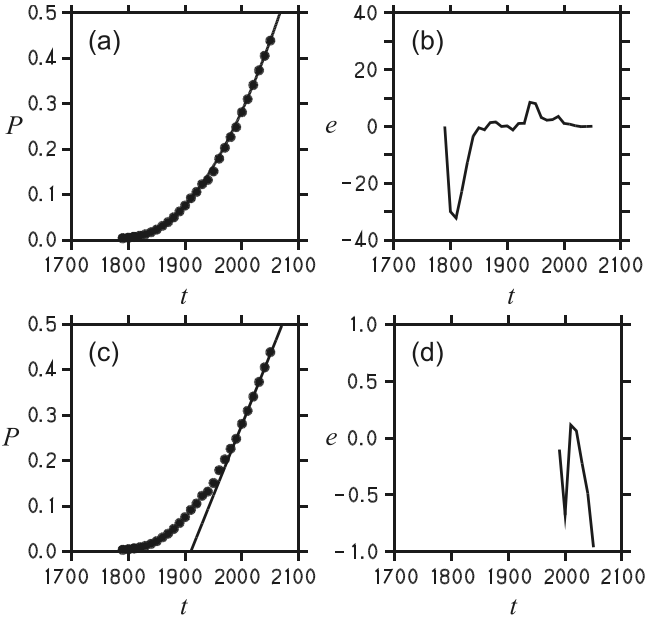



**Fig. 1.12.** Polynomial models for the development of the U.S. population $P$ (in $10^9$) in time $t$. The *dots* represent the U.S. population data. The *line* in (a) shows the quadratic polynomial (1.24), and (b) shows the corresponding model error in %. The linear model (1.25) and its error are shown in (c) and (d), respectively.

**Linear Polynomial.** Another approach is to focus the modeling on the data range that is most relevant: the population data for $t \geq 1990$. Figure 1.12 shows that a linear model may work for this data range. From Eq. (1.11) combined with the population data at 2000 and 2040 given in Table 1.5 we obtain the function

$$P = 0.28 + \frac{t - 2000}{322} = \frac{t - 1910}{322}. \tag{1.25}$$

The model error is shown in Fig. 1.12d for the range $t \geq 1990$. The magnitude of the relative error is smaller than 1%, which means that Eq. (1.25) represents a very accurate model.

### *1.4.3 Model Evaluation*

Four population models were presented in Sects. 1.4.1 and 1.4.2 – but which model is the best one? Let us consider the properties of reasonable and optimal models to prepare the answer to this question.

**Table 1.6** Basic properties of reasonable and optimal mathematical models.

| Basic Properties of a Reasonable Model | Basic Properties of an Optimal Model |
|---|---|
| 1a. Agreement with the trend of observations. | 1a. Accurate agreement with observations. |
| 1b. Agreement with observations over a certain data range of one case. | 1b. Accurate agreement with all observations of one class of problems. |
| 2a. Information in addition to observations results from using a function. | 2a. Information in addition to observations results from a model concept. |
| 2b. Interpolation and extrapolation of information given by observations. | 2b. Valuable new insight into the mechanisms of observed processes. |

**Reasonable and Optimal Models.** Characteristic properties of reasonable and optimal models are summarized in Table 1.6. There are also other criteria like the simplicity of formulation, cost and ease of use, but these additional criteria are not really relevant to the problems considered here. A basic requirement for a model is given by support through observations. The agreement with observations can range from a reasonable agreement with the trend of observations (reasonable model) up to an accurate agreement (errors below 1%) with observations (optimal model). An essential difference between models is given by the range over which a model agrees with observations. A reasonable model should be applicable to a certain data range of one case, whereas an optimal model should be applicable to all observations of relatively similar cases. Models also differ by the kind of support for the information provided in addition to the information given by observations, i.e., the basis for the modeling approach. Such support can range from the use of a function as given for polynomial models up to the use of a modeling concept that explains the nature of a class of problems (optimal model). In the end, the most basic requirement for a model is that the model has to be helpful, which means a model should provide more information than given by the observations. Such additional information can range from the interpolation and extrapolation of information given by observations (reasonable model) up to a valuable new insight into the mechanisms of observed processes (optimal model).

**Linearized Conceptual Versus Polynomial Modeling.** Let us use the criteria for reasonable and optimal models described in the preceding paragraph to evaluate the standard of linearized conceptual and polynomial models presented in Sects. 1.4.1 and 1.4.2, respectively. All the four population models satisfy the criteria for a reasonable model, but there are differences regarding the properties of optimal models. The advantage of linearized conceptual models is that these models are based on modeling concepts. The disadvantage of these two models is that these concepts are supported by observations only for a certain range of data values, and there are two models that have the same type of support. Do these models provide valuable new insight into the mechanism of this process? The

latter is not the case because there is no convincing evidence for the concepts represented by these two models. The advantage of polynomial models is their range of applicability and accuracy: we have the choice between a rather accurate quadratic model for the entire data range and a simple and very accurate linear model that describes the development for $t \geq 1990$ with a relative error below 1%. The disadvantage of polynomial models is that they do not present an explanation for the mechanism of this process: such models describe what we see anyway. Thus, no single out of the four population models considered represents an optimal model. How is it possible to develop an optimal model? Polynomial models do not have the potential to represent an optimal model (polynomial models are case dependent so that their model parameters change for every case considered). Thus, the only way is to refine the modeling concepts considered above to obtain a model that agrees well with observations over the entire data range.

## 1.5 The Advantage of Modeling: Global Warming Modeling

So far, we discussed two approaches for the modeling of given observations, and we compared characteristic features of the modeling approaches considered. These modeling approaches provide functions that agree in some way with the observed data. Such functions are helpful because they extend the information given by the observations via the interpolation and extrapolation of data. On the other hand, such functions do not directly provide valuable new insight into the nature of observed trends. The latter may require combinations of conclusions obtained from models. Such a combination of analytical results that are derived from observations will be described here in conjunction with the discussion of a relevant problem: the modeling of global warming.

### *1.5.1 The Greenhouse Effect*

**Greenhouse Effect.** The greenhouse effect is the process by which radiative energy (heat) leaving the Earth's surface is absorbed by certain atmospheric gases – water vapor ($H_2O$) is the most abundant greenhouse gas, followed by carbon dioxide ($CO_2$) and other trace gases – called greenhouse gases. The greenhouse gases re-radiate this energy in all directions, including back down towards the Earth's surface. Without the natural greenhouse effect, the Earth's temperature would be about −18°C instead of its present +14°C. Thus, the greenhouse effect helps to regulate the temperature of our planet: it is essential for life on Earth.

**Table 1.7** HadCRUT3 global temperature anomaly data (consisting of annual differences from 1961–1990 normals) in ºC from 1850–2008.

| $t$ | $T$ | $t$ | $T$ | $t$ | $T$ | $t$ | $T$ | $t$ | $T$ |
|---|---|---|---|---|---|---|---|---|---|
| 1850 | –0.402 | 1882 | –0.211 | 1914 | –0.322 | 1946 | –0.221 | 1978 | –0.049 |
| 1851 | –0.315 | 1883 | –0.242 | 1915 | –0.225 | 1947 | –0.202 | 1979 | 0.056 |
| 1852 | –0.345 | 1884 | –0.325 | 1916 | –0.461 | 1948 | –0.245 | 1980 | 0.102 |
| 1853 | –0.368 | 1885 | –0.340 | 1917 | –0.538 | 1949 | –0.230 | 1981 | 0.130 |
| 1854 | –0.321 | 1886 | –0.261 | 1918 | –0.354 | 1950 | –0.339 | 1982 | 0.008 |
| 1855 | –0.291 | 1887 | –0.330 | 1919 | –0.387 | 1951 | –0.172 | 1983 | 0.187 |
| 1856 | –0.438 | 1888 | –0.249 | 1920 | –0.323 | 1952 | –0.088 | 1984 | –0.011 |
| 1857 | –0.491 | 1889 | –0.150 | 1921 | –0.287 | 1953 | –0.055 | 1985 | –0.018 |
| 1858 | –0.529 | 1890 | –0.378 | 1922 | –0.374 | 1954 | –0.230 | 1986 | 0.022 |
| 1859 | –0.424 | 1891 | –0.332 | 1923 | –0.335 | 1955 | –0.291 | 1987 | 0.167 |
| 1860 | –0.378 | 1892 | –0.381 | 1924 | –0.364 | 1956 | –0.326 | 1988 | 0.163 |
| 1861 | –0.568 | 1893 | –0.453 | 1925 | –0.279 | 1957 | –0.089 | 1989 | 0.096 |
| 1862 | –0.616 | 1894 | –0.391 | 1926 | –0.184 | 1958 | –0.023 | 1990 | 0.248 |
| 1863 | –0.386 | 1895 | –0.313 | 1927 | –0.253 | 1959 | –0.114 | 1991 | 0.197 |
| 1864 | –0.434 | 1896 | –0.150 | 1928 | –0.240 | 1960 | –0.137 | 1992 | 0.055 |
| 1865 | –0.295 | 1897 | –0.185 | 1929 | –0.372 | 1961 | –0.044 | 1993 | 0.102 |
| 1866 | –0.287 | 1898 | –0.386 | 1930 | –0.173 | 1962 | –0.033 | 1994 | 0.163 |
| 1867 | –0.336 | 1899 | –0.235 | 1931 | –0.141 | 1963 | –0.047 | 1995 | 0.276 |
| 1868 | –0.160 | 1900 | –0.142 | 1932 | –0.177 | 1964 | –0.315 | 1996 | 0.123 |
| 1869 | –0.264 | 1901 | –0.257 | 1933 | –0.349 | 1965 | –0.219 | 1997 | 0.355 |
| 1870 | –0.257 | 1902 | –0.384 | 1934 | –0.199 | 1966 | –0.157 | 1998 | 0.515 |
| 1871 | –0.265 | 1903 | –0.477 | 1935 | –0.203 | 1967 | –0.151 | 1999 | 0.262 |
| 1872 | –0.262 | 1904 | –0.501 | 1936 | –0.161 | 1968 | –0.129 | 2000 | 0.238 |
| 1873 | –0.296 | 1905 | –0.321 | 1937 | –0.046 | 1969 | –0.004 | 2001 | 0.400 |
| 1874 | –0.412 | 1906 | –0.281 | 1938 | –0.028 | 1970 | –0.072 | 2002 | 0.455 |
| 1875 | –0.328 | 1907 | –0.446 | 1939 | 0.008 | 1971 | –0.178 | 2003 | 0.457 |
| 1876 | –0.330 | 1908 | –0.529 | 1940 | 0.013 | 1972 | –0.054 | 2004 | 0.432 |
| 1877 | 0.015 | 1909 | –0.564 | 1941 | 0.066 | 1973 | 0.060 | 2005 | 0.479 |
| 1878 | 0.031 | 1910 | –0.559 | 1942 | –0.055 | 1974 | –0.226 | 2006 | 0.422 |
| 1879 | –0.222 | 1911 | –0.566 | 1943 | –0.038 | 1975 | –0.134 | 2007 | 0.404 |
| 1880 | –0.210 | 1912 | –0.476 | 1944 | 0.087 | 1976 | –0.229 | 2008 | 0.296 |
| 1881 | –0.229 | 1913 | –0.463 | 1945 | –0.030 | 1977 | 0.059 | | |

**Table 1.8** $CO_2$ concentrations in ppmv from 1959–2007. The data were measured at the Mauna Loa Observatory in Hawaii (ftp://ftp.cmdl.noaa.gov/ccg/co2/trends/co2_annmean_ mlo.txt).

| $t$ | $CO_2$ | $t$ | $CO_2$ | $t$ | $CO_2$ | $t$ | $CO_2$ | $t$ | $CO_2$ |
|---|---|---|---|---|---|---|---|---|---|
| 1959 | 315.98 | 1969 | 324.62 | 1979 | 336.78 | 1989 | 352.90 | 1999 | 368.14 |
| 1960 | 316.91 | 1970 | 325.68 | 1980 | 338.68 | 1990 | 354.16 | 2000 | 369.40 |
| 1961 | 317.64 | 1971 | 326.32 | 1981 | 340.11 | 1991 | 355.48 | 2001 | 371.07 |
| 1962 | 318.45 | 1972 | 327.45 | 1982 | 341.22 | 1992 | 356.27 | 2002 | 373.17 |
| 1963 | 318.99 | 1973 | 329.68 | 1983 | 342.84 | 1993 | 356.95 | 2003 | 375.78 |
| 1964 | 319.62 | 1974 | 330.17 | 1984 | 344.41 | 1994 | 358.64 | 2004 | 377.52 |
| 1965 | 320.04 | 1975 | 331.08 | 1985 | 345.87 | 1995 | 360.62 | 2005 | 379.76 |
| 1966 | 321.38 | 1976 | 332.05 | 1986 | 347.19 | 1996 | 362.36 | 2006 | 381.85 |
| 1967 | 322.16 | 1977 | 333.78 | 1987 | 348.98 | 1997 | 363.47 | 2007 | 383.71 |
| 1968 | 323.04 | 1978 | 335.41 | 1988 | 351.45 | 1998 | 366.50 | | |

**Global Warming.** Observations reveal a significant increase of Earth's global surface temperature after about 1980. This development is very dangerous. An increase in global temperature may cause glacial retreat, Arctic shrinkage, and worldwide sea level rise. Changes in the amount and pattern of precipitation may result in flooding and drought. There also may be changes in the frequency and intensity of extreme weather events. The reason for the current global warming is given by human activity leading to an enhancement of the greenhouse effect by the emission of greenhouse gases (mostly $CO_2$ from combustion of coal, oil, and gas; plus a few other trace gases) through fossil fuel combustion and deforestation: pre-industrial levels of $CO_2$ (prior to the start of the Industrial Revolution) were about 280 parts per million by volume (ppmv), and current levels are greater than 380 ppmv. The global concentration of $CO_2$ in our atmosphere today far exceeds the natural range over the last 650,000 years of 180 to 300 ppmv.

**Temperature Increase.** Evidence for the global warming of the Earth is given by measurements of the global temperature anomaly. The global temperature anomaly represents the deviation from a certain standard temperature. Usually, the standard temperature is based on averaged temperature values over 30 years. Here, we consider averaged temperatures from 1961–1990. It is relevant to note that these standard temperature values depend on the location of measurements. Table 1.7 shows global temperature anomaly data. In particular, this table shows the HadCRUT3 annual global surface temperature anomaly data in ºC from 1850 to 2008 reported by the UK Met Office Hadley Centre for Climate Change (Rayner et al. 2003, Brohan et al. 2006, http://hadobs.metoffice.com, http://www.metoffice.gov.uk/research/hadleycentre/obsdata/HadCRUT3.html).

**$CO_2$ Emission Increase.** Evidence for the increasing concentration of atmospheric $CO_2$ is provided by the NOAA Earth System Research Laboratory (ESRL) data of atmospheric $CO_2$ concentrations in ppmv from 1959–2007 (http://www.esrl.noaa.gov/gmd/ccgg/trends, Tans 2008). The data were measured (remote from local sources of pollution) at the Mauna Loa Observatory in Hawaii at an altitude of about 4 km on the peak of the Mauna Loa mountain in Hawaii (data given in Table 1.8).

### *1.5.2 $CO_2$ Concentration and Global Temperature Modeling*

**$CO_2$ Concentration Modeling.** First, let us try to develop a model for the $CO_2$ concentration development. The $CO_2$ development shown in Fig. 1.13a indicates relatively little deviations from a linear function. Thus, we consider the function

$$CO_2 = CO_2(t_0) + S(t)(t - t_0) \tag{1.26}$$

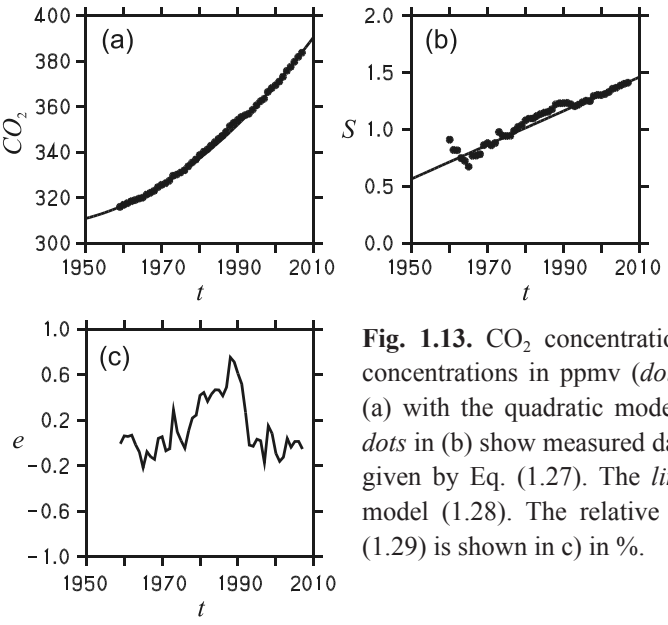



**Fig. 1.13.** $CO_2$ concentrations. Measured $CO_2$ concentrations in ppmv (*dots*) are compared in (a) with the quadratic model (1.29) (*line*). The *dots* in (b) show measured data for $S(t)$, which is given by Eq. (1.27). The *line* shows the linear model (1.28). The relative error of the model (1.29) is shown in c) in %.

as a basic model for the $CO_2$ development, with initial values $t_0 = 1959$ and $CO_2(t_0) = 316$ ppmv. The slope function $S(t)$ is defined by the rewritten Eq. (1.26),

$$S(t) = \frac{CO_2 - CO_2(t_0)}{t - t_0}, \tag{1.27}$$

which enables the calculation of $S(t)$ from data. Figure 1.13a shows that $S(t)$ can be modeled by the following linear function of $t$,

$$S(t) = 0.7\left(1 + \frac{t - t_0}{47}\right). \tag{1.28}$$

The combination of Eq. (1.26) with this expression for $S(t)$ leads then to a quadratic model for the development of the $CO_2$ concentration:

$$CO_2 = 316 + 0.7\left(1 + \frac{t - 1959}{47}\right)(t - 1959). \tag{1.29}$$

The small deviations seen in Fig. 1.13a do not affect the model performance. These deviations are multiplied with ($t$–1959), which is small in this data range. Figure 1.13c indicates the relative error magnitude as being smaller than 0.8%.

**Global Temperature Modeling.** The modeling of the global temperature increase represents a rather challenging problem (see Fig. 1.14a). It is obvious that the global temperature data are significantly affected by randomness. To get a guideline for dealing with this randomness, let us consider averaged temperatures.

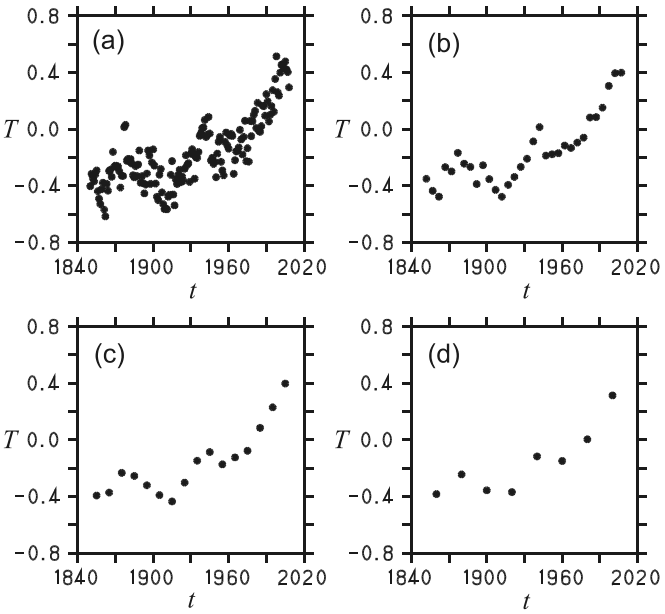



**Fig. 1.14.** Global temperature anomaly data: (a) temperature data without any averaging; (b), (c), and (d) show the temperature data averaged over 5, 10, and 20 years, respectively.

Such averaged temperature values are shown in Fig. 1.14, where averages over 5, 10, and 20 years are applied, respectively. For example, the 5-year average at $t = 1852$ is obtained by adding the temperatures from 1850 to 1854 and dividing the sum by five. Figure 1.14d demonstrates that the amount of randomness can be reduced in this way. To derive a model for the global temperature anomaly we will use the relatively smooth 20-year averaged data shown in Fig. 1.14d. The data support the view that the temperature values are approximately constant up to about $t = 1920$. We may assume that a constant value $T = -0.35$ works for this range. For later times, the temperature values seem to increase like a power function. So let's try to work with the model function

$$T = -0.35 + \left(\frac{t-1840}{a}\right)^b . \tag{1.30}$$

The reference to 1840 corresponds to the assumption that $T = -0.35$ for $t$ values that are not too far from 1840. The next step is to determine the model parameters $a$ and $b$. For doing this, it is convenient to linearize Eq. (1.30),

$$\ln(T + 0.35) = b \ln\left(\frac{t-1840}{a}\right). \tag{1.31}$$

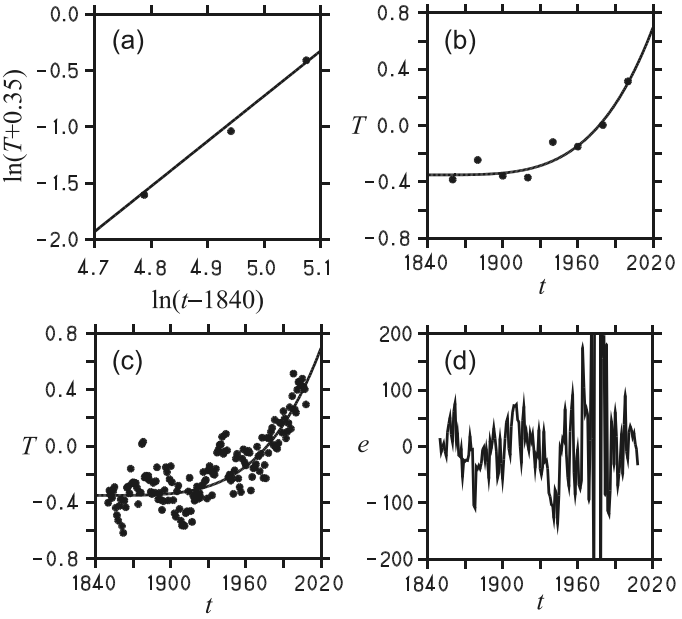



**Fig. 1.15.** The global temperature increase. The *dots* in (a) show $\ln(T + 0.35)$ versus $\ln(t - 1840)$ according to the 20-year averaged observations at 1960, 1980, and 2000. The *line* presents Eq. (1.32). A comparison between 20-year averaged observations and Eq. (1.33) is shown in (b). Fig. (c) compares the model (1.33) with the non-averaged temperature anomaly data. Fig. (d) shows the relative error $e$ (in %) of the model (1.33).

A plot of $\ln(T + 0.35)$ versus $\ln(t - 1840)$ may show whether the power function approach applied will work. However, Fig. 1.14d shows that there are three data values that are smaller than $T = -0.35$, which means that the natural logarithm of $T + 0.35$ would have negative arguments. First of all, the parameter $a$ and $b$ values have to describe correctly the increase of temperature values for $t \geq 1960$. Thus, we will only apply the averaged data at 1960, 1980, and 2000. A corresponding plot is given in Fig. 1.15a. This figure also shows that the function

$$\ln(T + 0.35) = 4 \ln\left(\frac{t - 1840}{178}\right) \tag{1.32}$$

does well agree with the observed data. The use of a slope that is slightly higher than the value four applied here may result in a slightly improved agreement between the model and observations. However, it would be more difficult to illustrate the relationship between the global temperature anomaly and $CO_2$ in this case (see Sect. 1.5.3). Apart from that, it is relevant to see that the 20-year averages are only used as guideline for the model development. According to

Eq. (1.32), the global temperature anomaly development is described by

$$T = -0.35 + \left( \frac{t - 1840}{178} \right)^4 . \tag{1.33}$$

Figures 1.15b and 1.15c reveal that this model agrees reasonably well with both the 20-year averaged observations and the entire range of observations.

**Global Temperature Model Evaluation.** Figure 1.15d shows the relative error for the model (1.33) – the relative error is large in general. In particular, the maximum error magnitude is given by 6340%. A closer look at the data shows that huge error values of more than ±200% appear in an area where the magnitude of $T$ is relatively small, $-0.06 \le T \le 0.06$ (between 1971 and 1982). For example, at $t = 1977$, the difference between the observed and measured temperature anomaly is given by 0.058ºC. However, by dividing this difference by the small modeled temperature of $T = 0.000915$ºC one ends up with a huge relative error: $e = 63.4$, this means 6340%. The error calculation in this range of small modeled temperature anomaly values is simply inappropriate. This means such relative error values should not be considered. The typical error values fluctuate in the range ±100%. These error values pose questions about the suitability of evaluating the performance of models for observations that involve a significant amount of randomness. Such questions will be addressed in Chap. 2.

### *1.5.3 The Advantage of Modeling*

**Model Applications.** Compared to the measured data, one advantage of the $CO_2$ concentration formula (1.29) and the global temperature anomaly formula (1.33) is that we can use these models for making predictions. According to Eq. (1.29) we find a $CO_2$ concentration of 710.8 ppmv by the end of the 21st century (at $t = 2100$). This value is 154% above the pre-industrial concentration value of 280 ppmv! This formula also shows that the $CO_2$ emission per year is increasing. In 2000, the $CO_2$ concentration increase rate (the difference of $CO_2$ emissions from 2000 to 2001) was 2 ppmv $yr^{-1}$. In 2010, the increase rate has been 2.3 ppmv $yr^{-1}$. According to Eq. (1.33), we have to expect a global temperature anomaly of 4.2ºC by the end of the 21st century, this means a temperature increase of 4.55ºC compared to the temperatures in 1840! Obviously, such a temperature increase will imply dramatic consequences. These predictions are helpful for the understanding of the dimension of this problem, but they do not explicitly explain the relationship between the observed increase of the global temperature anomaly and the $CO_2$ concentration, which is relevant to deal with this problem in an appropriate way.

***T*-$CO_2$ Relation.** To clearly see the effect of increasing $CO_2$ concentrations on the temperature increase, let us combine Eqs. (1.33) and (1.29). The time $t$ can be expressed as a function of $T$ according to Eq. (1.33),

$$t = 1840 + 178(T + 0.35)^{1/4}. \tag{1.34}$$

By relating $t$ to $T$ in Eq. (1.29) we find

$$\begin{aligned} CO_2 &= 316 + 0.7\left(1 + \frac{178(T+0.35)^{1/4} - 119}{47}\right)\left(178(T+0.35)^{1/4} - 119\right) \\ &= 316 + \frac{0.7}{47}178^2\left(\frac{47}{178} + \left((T+0.35)^{1/4} - \frac{119}{178}\right)\right)\left((T+0.35)^{1/4} - \frac{119}{178}\right). \end{aligned} \tag{1.35}$$

This relation can be written as a quadratic equation:

$$0 = \left((T+0.35)^{1/4} - \frac{119}{178}\right)^2 + \frac{47}{178}\left((T+0.35)^{1/4} - \frac{119}{178}\right) - \frac{47}{0.7}\frac{CO_2 - 316}{178^2}. \tag{1.36}$$

The solution of this equation reads

$$\begin{aligned} T &= -0.35 + \left[\frac{119}{178} - \frac{47}{356} \pm \sqrt{\frac{1}{4}\frac{47^2}{178^2} + \frac{47}{0.7}\frac{CO_2 - 316}{178^2}}\right]^4 \\ &= -0.35 + \left[\frac{119}{178} + \frac{47}{356}\left(\mp\sqrt{1 + \frac{4}{0.7 \cdot 47}(CO_2 - 316)} - 1\right)\right]^4. \end{aligned} \tag{1.37}$$

The square root combined with the positive sign provides for $CO_2$ = 316 ppmv at $t$ = 1959 a global temperature anomaly $T$ = –0.35 + (119/178)$^{1/4}$ = –0.15, which agrees with the consequence of Eq. (1.33). Thus, the formula for $T(CO_2)$ reads

$$T = -0.35 + \left[\frac{119}{178} + \frac{47}{356}\left(\sqrt{1 + \frac{40}{329}(CO_2 - 316)} - 1\right)\right]^4. \tag{1.38}$$

A plot of temperature variations according to Eq. (1.38) is shown in Fig. 1.16. The range of $CO_2$ values considered corresponds to the $CO_2$ variation between 1959 and 2010 (Eq. (1.29) predicts a value of $CO_2$ = 390.4 ppmv at $t$ = 2010). Figure 1.16 reveals an almost linear increase of temperature values with the $CO_2$ concentrations.

**Linear *T*-$CO_2$ Relation.** To see the difference between the $T$-$CO_2$ relation (1.38) and a linear function, let us derive a linear approximation to this relation. The first-order Taylor series expansion of Eq. (1.38) at any reference time $t_0$ reads

$$T^{(LIN)} = T(t_0) + \left(\frac{dT}{dCO_2}\right)_{CO_2 = CO_2(t_0)} \left(CO_2 - CO_2(t_0)\right). \tag{1.39}$$

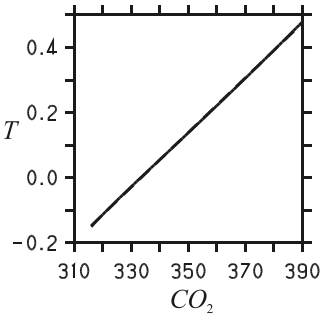



**Fig. 1.16.** The temperature-$CO_2$ relation. The *solid line* shows this relation according to Eq. (1.38). The *dashed line* shows the linear approximation (1.41).

The derivative that appears here is given by

$$\frac{dT}{dCO_2} = 4(T+0.35)^{3/4}\,\frac{47}{356}\,\frac{40/329}{2\sqrt{1+40(CO_2-316)/329}}$$
$$= \frac{47^2}{356^2}\,\frac{80}{329}\,\frac{(T+0.35)^{3/4}}{(T+0.35)^{1/4}-119/178+47/356}$$
$$= \frac{47^2}{16\cdot 89^2}\,\frac{16\cdot 5}{7\cdot 47}\,\frac{(T+0.35)^{3/4}}{(T+0.35)^{1/4}-191/356} = \frac{235}{7\cdot 89^2}\,\frac{(T+0.35)^{3/4}}{(T+0.35)^{1/4}-191/356}. \tag{1.40}$$

Here, Eq. (1.38) is used to replace $CO_2$ by $T$ in order to obtain simpler expressions. Different linear approximations can be found in dependence on the choice of $t_0$. Here, we will use $t_0 = 1976.9$ because $T(t_0) = 0$ for this case. The calculation of $T$, $CO_2$, and $dT/dCO_2$ at $t_0 = 1976.9$ leads then to the following linear approximation to Eq. (1.38),

$$T^{(LIN)} = \frac{CO_2 - 333.3}{120.6}. \tag{1.41}$$

Figure 1.16 shows that this linear function agrees very well with the exact relation (1.38): there is hardly any observable difference. This formula shows that the global temperature anomaly grows linearly with the $CO_2$ concentration. Thus, the increase of the global temperature can be reduced by reducing the $CO_2$ emissions. Such a clear explanation regarding the driving force of global warming cannot be obtained by means of observations only, which illustrates the value of modeling. The relevance of the numbers involved in Eq. (1.41) can be seen by relating $T^{(LIN)}$ to a nondimensional $CO_2$ change (we have $CO_2(1976.9) = 333.3$ ppmv),

$$T^{(LIN)} = 2.76\frac{CO_2 - CO_2(1976.9)}{CO_2(1976.9)}. \tag{1.42}$$

Hence, a 10% increase of the $CO_2$ concentration compared to the level of 1976.9 resulted in a temperature increase of 0.28ºC.

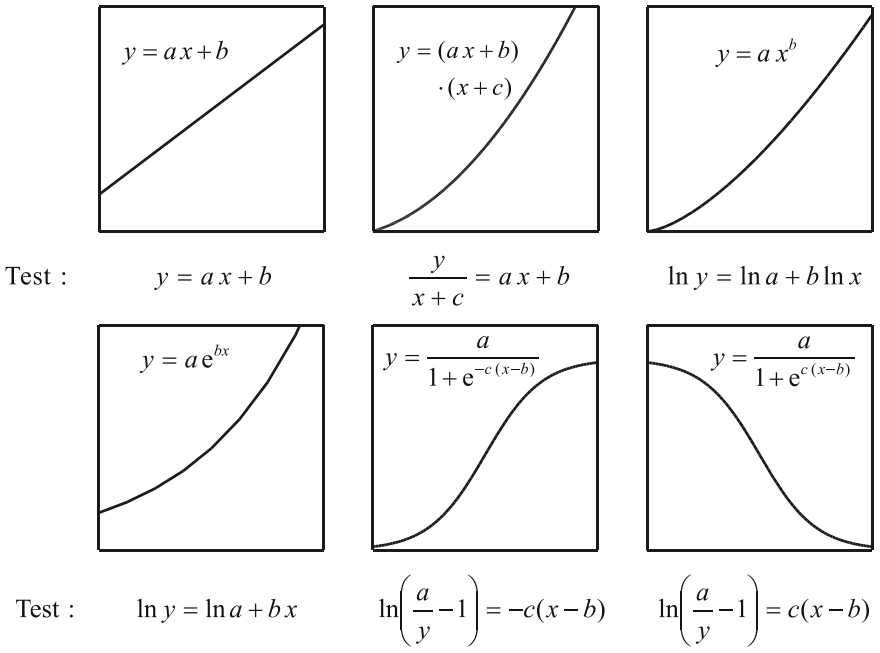



**Fig. 1.17.** Basic function types that can be linearized. The linearity test for the applicability of the each model type is shown below the pictures.

## 1.6 Summary

Let us summarize the observations made in this chapter by following the basic four steps of the modeling process described in the beginning.

**Data Transformations.** The most efficient way to reduce modeling problems is to apply data transformations such that linear relations between redefined data can be considered. The use of this approach was discussed here in terms of several examples: the modeling of the energy consumption, planetary motion, the total vehicular stopping distance, the modeling of population dynamics, and the global warming. Figure 1.17 illustrates that a variety of data trends can be described by basic functions that can be linearized in terms of redefined data. For all these functions, Figure 1.17 also shows the linearity test (these linearity relations have to be satisfied in order to use the corresponding model type). The consideration of such linear relations is also the most convenient way for using observations for the calculation of optimal model parameters (which minimize the deviations between the model and observations: see Chap. 2).

**Model Development.** Apart from the use of linear models (if required for redefined data), the simplest modeling approach is the use of polynomial models. One way to work with polynomials is the use of exact polynomial models. This modeling approach appears to be very attractive because exact polynomial models agree exactly with any data set considered, and they are completely defined for any given set of data (which minimizes the modeling effort). However, it turns out that polynomials of higher than third order are often not helpful for the solution of practical questions. Such polynomials reveal an extreme sensitivity to small data variations, which often implies severe oscillations of the model between data points and a completely incorrect behavior at the end of the interval considered. These problems can be avoided by the use of low-order polynomials (linear, quadratic, or cubic functions). Their use requires, however, a careful choice of the data points that determine the polynomial considered. Examples for the usefulness of this approach were given here regarding the discussion of population dynamics.

**Model Evaluation.** The next step after developing a model is its evaluation. Table 1.6 summarizes criteria for the evaluation of the model types considered here (the agreement with observed data and the information provided by models). The basic goal of such a model evaluation is to provide evidence that the proposed model fulfills the standards of a reasonable model: the model will be helpful in this case. Regarding the population dynamics it was shown that it is well possible to develop several reasonable models for any problem considered, but such models do often not represent optimal models. A relevant requirement for the development of an optimal model is the use of a concept for the model development. The development of model concepts will be discussed in detail in the following chapters. An example for an optimal model was given here by the highly accurate Kepler's Third Law (the model concept used in this case is the linearity between $\ln T_P$ and $\ln r$, which is supported by theory: see Sect. 3.3.2).

**Demonstration of the Advantage of Models.** The purpose of the development of a mathematical model is to obtain more information about the state or the process considered. Usually, the development of models that have support from data is helpful due to the interpolation and extrapolation of data values, which is obtained by the model. However, models can also provide valuable insight in addition to the information provided by observations. An example for such a significant contribution of a model was given here regarding the discussion of global warming. We derived the linear temperature-$CO_2$ relation (1.41), which explains the driving mechanism of this process. Such a demonstration of the advantages of mathematical models does not directly follow from the development of models in agreement with observations: appropriate combinations of model results may be needed to derive conclusions that are as simple and clear as possible.

## 1.7 Exercises

**1.2.1** Consider the following data.

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $y$ | 100 | 25 | 11 | 6 | 4 | 3 | 2 |

a) Plot $\ln y$ versus $x$ and $\ln x$. Compare the data with a linear function if one of these plots supports the use of a linear function.
b) Graph the table data and the model that follows from the above relation.
c) Calculate the relative error of your model in %.
d) At which $x$ is the value of $y$ at $x = 2.5$ reduced by 25%?

**1.2.2** Consider the following data.

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $y$ | 2 | 35 | 150 | 500 | 1250 | 2500 | 5000 |

a) Plot $\ln y$ versus $x$ and $\ln x$. Compare the data with a linear function if one of these plots supports the use of a linear function.
b) Graph the table data and the model that follows from the above relation.
c) Calculate the relative error of your model in %.
d) At which $x$ is the value of $y$ at $x = 2.5$ increased by 150%?

**1.2.3** Consider the following data.

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $y$ | 460 | 280 | 170 | 103 | 63 | 38 | 23 |

a) Plot $\ln y$ versus $x$ and $\ln x$. Compare the data with a linear function if one of these plots supports the use of a linear function.
b) Graph the table data and the model that follows from the above relation.
c) Calculate the relative error of your model in %.
d) At which $x$ is the value of $y$ equal to $y(x = 0) / 100$?

**1.2.4** Consider the following data. Assume that the data set can be described by a function $y = c + a x^b$, where $a$, $b$, and $c$ are any constants.

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $y$ | 100 | 101 | 111 | 140 | 217 |

a) Linearize $y = c + a x^b$ and use the data to test the suitability of this assumption. Find a linear function for this data plot.
b) Graph the table data and the model that follows from the above relation.
c) Calculate the relative error of your model in %.
d) Find the $x$ at which the value of $y(x = 0)$ is doubled.

**1.2.5** Consider the following data.

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $y$ | −30 | −10 | −13 | −20 | −25 |

Assume that the data set can be described by a function $y = c + a\,x\,e^{bx}$, where $a$, $b$, and $c$ are any constants.

a) Linearize $y = c + a\,x\,e^{bx}$ and use the data to test the suitability of this assumption. Find a linear function for this data plot.

b) Graph the table data and the model that follows from the above relation.

c) Calculate the relative error of your model in %.

d) Calculate the maximum of $y(x)$.

**1.2.6** Consider the case where you have data that can be modeled by the function $y = y_0 + a\,(\exp[b(x–x_0)] - 1)$. Here, $a$, $b$, and $(x_0, y_0)$ are any constants. Explain how the data can be used to determine the constants.

**1.3.1** The U.S. Bureau of Public Roads determined the following total stopping distances $D$ (in ft) depending on the velocity $v$ (in mph) of cars.

| $v$ | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|
| $D$ | 42 | 73.5 | 116 | 173 | 248 | 343 | 464 |

a) Use the data to plot $\ln D$ versus $v$ and $\ln v$. Compare the data with linear functions that reveal the parameters of corresponding exponential and power function models.

b) Use the data to plot $D/v$. Compare the data with a linear and a quadratic function.

c) Plot $D$ according to the original data in comparison to the two models for $D/v$. Calculate the relative error of the two polynomial models.

d) Discuss the suitability of the models obtained for $D$. Identify one model that provides a formula that can be used to calculate the total stopping distance without using a calculator. Illustrate the use of this formula by three examples.

**1.3.2** Consider again the total stopping distances data given in problem 1.3.1.

a) Use the entire data set to define an exact polynomial of sixth order. Graph this polynomial and the data. Comment on the suitability of this model.

b) Find a way to improve the performance of this model significantly (you may replace one data point by other reasonable data).

c) Use three data points to define a reasonable quadratic polynomial.

d) Reduce the quadratic polynomial to a simple formula for $D$.

**1.3.3** Consider the development of the world population in time from 1804–2050 according to the Decennial Censuses, U.S. Census Bureau, U.S. Dept. of Commerce (World Almanac 2010). The population $P$ is measured in $10^9$ and $t$ refers to the year. The last two population values are projections.

| $t$ | 1804 | 1927 | 1960 | 1974 | 1987 | 1999 | 2009 | 2025 | 2050 |
|---|---|---|---|---|---|---|---|---|---|
| $P$ | 1.0 | 2.0 | 3.0 | 4.0 | 5.0 | 6.0 | 6.77 | 7.95 | 9.32 |

a) Use the data from 1804 to 2009 to define an exact polynomial of sixth order. Graph this polynomial and the data. Comment on the suitability of this model.
b) Use the data from 1960 to 2009 to define an exact polynomial of fourth order. Graph this polynomial and the data. Comment on the suitability of this model.
c) Use the data at 1960, 1987, and 2009 to define a polynomial of second order. Graph this polynomial and the data. Comment on the suitability of this model.
d) Use the 1960 and 2009 data to define a polynomial of first order. Graph this polynomial and the data. Comment on the suitability of this model.

**1.3.4** Consider the following atmospheric $CO_2$ concentrations (see Table 1.8).

| $t$ | 1995 | 1997 | 1999 | 2001 | 2003 | 2005 | 2007 |
|---|---|---|---|---|---|---|---|
| $CO_2$ | 360.62 | 363.47 | 368.14 | 371.07 | 375.78 | 379.76 | 383.71 |

a) Use all the data to define an exact polynomial of sixth order. Graph this polynomial and the data. Comment on the suitability of this model.
b) Use the data from 1995, 1999, 2003, and 2007 to define a polynomial of third order. Graph this polynomial and the data. Comment on the suitability of this model.
c) Use the data from 1995, 1999, and 2007 to define a polynomial of second order. Graph this polynomial and the data. Comment on the suitability of this model.
d) Use the 1995 and 2007 data to define a polynomial of first order. Graph this polynomial and the data. Comment on the suitability of this model.

**1.4.1** The following data describe the changes of a certain population $P$ in time $t$ (in days). Develop a model for the data based on the logistic model $P = a / [1 + b\, e^{c\, t})]$.

| $t$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $P$ | 43 | 58 | 76 | 90 | 97 | 106 | 112 |

a) Rewrite this function as a linear relation between redefined variables.
b) Graph the data such that the linearity assumption can be tested. Compare the data in this plot with a linear function to find the model parameters.
c) Present the model obtained and graph both the model and the original data given in the table. Graph the relative error of the model.
d) Use the model to predict the time at which the initial population at $t = 0$ is increased by a factor of 2.

**1.4.2** The following data describe the changes of a certain population $P$ in time $t$ (in days). Develop a model for the data based on the logistic model $P = a / [1 + b\,e^{c\,t})]$.

| $t$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $P$ | 96 | 92 | 83 | 69 | 50 | 31 | 17 | 8 |

a) Rewrite this function as a linear relation between redefined variables.
b) Graph the data such that the linearity assumption can be tested. Compare the data in this plot with a linear function to find the model parameters.
c) Present the model obtained and graph both the model and the original data given in the table. Graph the relative error of the model.
d) Determine the time at which $P = 1$.

**1.4.3** Consider again the world population data given in problem 1.3.2. Assume that the population $P$ can be described by the function

$$P = \frac{a}{1 + b\,e^{ct}} + d,$$

where $a$, $b$, $c$, and $d$ are any constants. For a certain time period before 1804, the population can be approximated by a constant value $P = 1$. Assume that the population density levels off finally at a value of $P = 11$.
a) Rewrite the model for $P$ as a linear relation between redefined variables.
b) Graph the data such that the linearity assumption can be tested. Compare the data in this plot with a linear function to find the model parameters.
c) Present the model obtained and graph both the model and the original data given in the table. Graph the relative error of the model.
d) Find the time at which the population change $dP / dt$ has a maximum.

**1.5.1** Consider the following data for the global temperature anomaly $T$ and atmospheric $CO_2$ concentration (see also Tables 1.7 and 1.8).
a) Use the data to graph $T$ as a function of $CO_2$. Graph in the same figure three linear functions. The first and second linear functions have to provide an upper and lower bound for $T$ as a function of $CO_2$. The third linear function, which presents the best model, has to show the average of the lower and the upper bound.

| $t$ | $CO_2$ | $T$ | $t$ | $CO_2$ | $T$ | $t$ | $CO_2$ | $T$ |
|---|---|---|---|---|---|---|---|---|
| 1961 | 317.64 | –0.044 | 1977 | 333.78 | 0.059 | 1993 | 356.95 | 0.102 |
| 1963 | 318.99 | –0.047 | 1979 | 336.78 | 0.056 | 1995 | 360.62 | 0.276 |
| 1965 | 320.04 | –0.219 | 1981 | 340.11 | 0.130 | 1997 | 363.47 | 0.355 |
| 1967 | 322.16 | –0.151 | 1983 | 342.84 | 0.187 | 1999 | 368.14 | 0.262 |
| 1969 | 324.62 | –0.004 | 1985 | 345.87 | –0.018 | 2001 | 371.07 | 0.400 |
| 1971 | 326.32 | –0.178 | 1987 | 348.98 | 0.167 | 2003 | 375.78 | 0.457 |
| 1973 | 329.68 | 0.060 | 1989 | 352.90 | 0.096 | 2005 | 379.76 | 0.479 |
| 1975 | 331.08 | –0.134 | 1991 | 355.48 | 0.197 | 2007 | 383.71 | 0.404 |

b) Present the three linear functions according to the formula

$$T = \frac{CO_2 - a}{b} \pm c,$$

where $a$, $b$, and $c$ are constants that you have to provide.

c) Discuss the advantages and disadvantages of this model.

d) Consider $T = 0.4$ºC and $CO_2 = 384$ ppmv at $t = 2007$. Assume that the $CO_2$ concentration increases by 10% in comparison to $CO_2 = 384$ ppmv. What will be the corresponding temperature increase compared to $T = 0.4$ºC? You can neglect the temperature uncertainty.

**1.5.2** Consider the following data for the global temperature anomaly $T$ and atmospheric $CO_2$ concentration (see also Tables 1.7 and 1.8).

| $t$ | 1995 | 1997 | 1999 | 2001 | 2003 | 2005 | 2007 |
|---|---|---|---|---|---|---|---|
| $CO_2$ | 360.62 | 363.47 | 368.14 | 371.07 | 375.78 | 379.76 | 383.71 |
| $T$ | 0.276 | 0.355 | 0.262 | 0.400 | 0.457 | 0.479 | 0.404 |

a) Use the data to graph $T$ and $CO_2$ as functions of $t$. Graph in the $CO_2$ figure a linear function that approximates the $CO_2$ data. Graph in the $T$ figure three linear functions. The first and second linear functions have to provide an upper and lower bound for $T$ as a function of $t$. The third linear function, which presents the best model, has to show the average of the lower and the upper bound.

b) Present the linear functions according to the formulas

$$CO_2 = a + b(t - 2007), \qquad T = \frac{t - 2007}{e} + c \pm d,$$

where $a$, $b$, $c$, $d$, and $e$ are any constants that you have to provide.

c) Combine $CO_2(t)$ and $T(t)$ to a function $T(CO_2)$ according to the formula

$$T = \frac{CO_2 - A}{B} \pm C,$$

where $A$, $B$, and $C$ are any constants that you have to provide.

d) Consider $T = 0.4$ºC and $CO_2 = 384$ ppmv at $t = 2007$. Assume that the $CO_2$ concentration increases by 10% in comparison to $CO_2 = 384$ ppmv. What will be the corresponding temperature increase compared to $T = 0.4$ºC? You can neglect the temperature uncertainty.

**1.5.3** Consider the same data as for problem 1.5.2. Use the $CO_2$ and $T$ data in the table to generate averaged $CO_2$ and $T$ data by averaging over three points. For example, the new $T(1997) = [T(1995) + T(1997) + T(1999)] / 3$, the new $T(1999) = [T(1997) + T(1999) + T(2001)] / 3$, and so on. You will find five new data values for $CO_2$ and $T$ in this way. Use the new data values to do a), b), c), and d) of problem 1.5.2.

# 2 Stochastic Analysis of Observations

The basic steps of developing a mathematical model for any given observations (e.g., a model for the global atmospheric temperature increase) were described in Chap. 1. In particular, the goal was to find a relatively simple analytical function (for example a linear function possibly formulated in terms of redefined variables) that is capable of representing the trend of observations, and to adjust then the parameters of this function such that the model agrees as good as possible with the observed data. The adjustment of model parameters was performed in this way on a purely empirical basis. Consequently, there is the question of how it is possible to optimize the performance of models by finding model parameters that provide an optimal agreement between the model and observations. This question can be addressed, first, by introducing and minimizing a model error that accounts for differences between the model and data, or, second, by using assumptions about the sort of randomness involved in observations. The use of the second approach requires knowledge regarding the description and modeling of randomness. This approach will be described in Chap. 10 in conjunction with the introduction of probability concepts for several variables. The first approach for finding optimal model parameters will be described in this chapter. The discussions of probability concepts in the following chapters are prepared in this way by demonstrating the relevance and application of some basic tools for the description of random data like the mean, variance, and correlation coefficient.

Section 2.1 explains the need for the development of optimal models. There is not only one optimization concept, but there are several concepts that can be used for the construction of optimal models. Essential advantages and disadvantages of different optimization concepts will be described in Sect. 2.2. The optimization of linear, quadratic, power and exponential function models will be addressed in the Sects. 2.3, 2.4, and 2.5, respectively. Section 2.6 deals with a summary of observations made in this chapter regarding the design of optimal models.

S. Heinz, *Mathematical Modeling*, DOI 10.1007/978-3-642-20311-4_2,

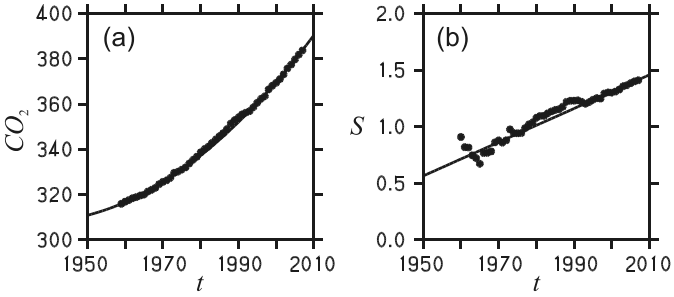



**Fig. 2.1.** The increase of the atmospheric $CO_2$ concentration in time $t$: (a) the $CO_2$ concentration in ppmv, and (b) the change $S = [CO_2 - CO_2(t_0)] / (t - t_0)$, where $t_0 =$ 1959 and $CO_2(t_0) = 316$ ppmv are used. The *dots* show measured data. The *lines* represent the models (1.28) and (1.29): see Chap. 1.

## 2.1 Motivation

**Atmospheric $CO_2$ Concentration.** Let us consider again the development of the atmospheric $CO_2$ concentration discussed in Sect. 1.5.2 to illustrate the need for the development of optimal models. The measured $CO_2$ concentration and the change $S = [CO_2 - CO_2(t_0)] / (t - t_0)$ of the $CO_2$ concentration in time $t$ are shown in Fig. 2.1 together with the models derived in Chap. 1. We see that the models are certainly appropriate, but there is no guarantee that the models applied are the best possible models, this means optimal models. The latter question is relevant regarding the importance of the global warming problem. It is also worth noting that there are not only a few problems (like the $CO_2$ concentration) that require the calculation of optimal model parameters. Instead, all the models developed on the basis of observations may benefit from an optimization. For example, Kepler's Third Law model derived in Sect. 1.2.2 agrees extremely well with observations, but there are also deviations between the model and observed data. Thus, the use of optimization techniques will be helpful for the improvement of the accuracy of the formulation of Kepler's Third Law (in particular for the calculation of the standard gravitational parameter). However, which concept can we use for optimizing the performance of models? One way to derive optimal model parameters would be the calculation of the relative error of several models and the use of the model that has a minimal relative error. However, the application of this approach is rather expensive: it requires the evaluation of a variety of models. There are also other approaches (see Sect. 2.2) that may be more appropriate. Thus, the question of the optimization of model parameters is a non-trivial question.

**Questions Considered.** The discussion in the preceding paragraph showed that there are several questions that have to be addressed:
- What is the best definition of a model error that has to be minimized?
- How difficult is the application of the best optimization concept?
- What is the advantage of using a concept for finding optimal model parameters?

The latter questions will be considered in the following sections of this chapter.

## 2.2 Model Errors

The minimizing of deviations between a model and observed data requires the definition of a global model error that becomes minimal. The term global means that we do not consider the local deviations at each data point, but we consider one error that represents a characteristic measure for all the local deviations. A closer look at this question shows that there are many ways to introduce a global error. Let us consider some characteristic advantages and disadvantages of several concepts for addressing this question.

### *2.2.1 Model Errors*

**Global Error Definitions.** A first possibility for defining a global error is to follow the idea of finding the model for which the largest local relative error is minimal. However, instead of considering normalized errors we define the global error by the maximal absolute value of all local deviations $Y_i - y_M(X_i)$ between given data $Y_i$ and the values of a model $y_M(x)$ at the positions $X_i$ considered,

$$E_C = \max_{i=1,N} \left| Y_i - y_M(X_i) \right|. \tag{2.1}$$

Here, $N$ refers to the number of data points considered. This global error is called the Chebyshev error. A second possibility is given by using the mean value of the absolute values of $Y_i - y_M(X_i)$,

$$E_A = \frac{1}{N} \sum_{i=1}^{N} \left| Y_i - y_M(X_i) \right|. \tag{2.2}$$

This error is called the least-absolute-deviations error $E_A$. A third possibility is to consider the mean value of squared deviations $Y_i - y_M(X_i)$,

$$E^2 = \frac{1}{N} \sum_{i=1}^{N} \left( Y_i - y_M(X_i) \right)^2. \tag{2.3}$$

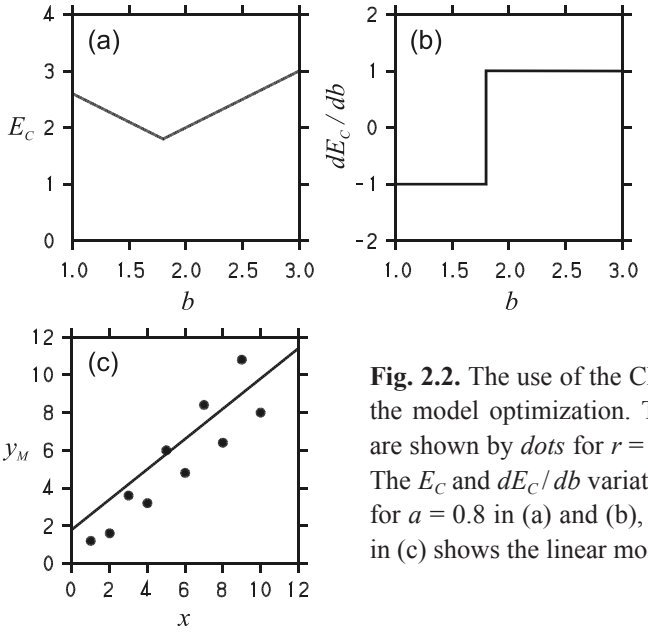



**Fig. 2.2.** The use of the Chebyshev criterion for the model optimization. The ($X_i$, $Y_i$) data (2.4) are shown by *dots* for $r = 0.2$ and $N = 10$ in (c). The $E_C$ and $dE_C / db$ variations with $b$ are shown for $a = 0.8$ in (a) and (b), respectively. The *line* in (c) shows the linear model (2.7).

This error is called the least-squares error. The consideration of $E^2$ refers to the fact that the mean value of squared deviations is obtained in this way. Obviously, there are many other possibilities to introduce global errors. For example, the second power in Eq. (2.3) can be replaced by any other power.

**Noise Model.** Which global error definition will be the best choice for finding optimal model parameters? To address the latter question we will apply the three global error definitions referred to above to the modeling of the following set of data values ($X_i$, $Y_i$),

$$X_i = i, \qquad Y_i = \left(1 - (-1)^i r\right) i. \tag{2.4}$$

Here, $i = 1, N$, where $N$ refers to the number of values ($X_i$, $Y_i$). For simplicity, we do only consider the case that $N$ is an even number. Equation (2.4) represents a simple model for noisy data. The parameter $r \geq 0$ characterizes the noise rate, for example $r = 0.2$ corresponds to 20% noise. The zero noise case is given by $X_i = Y_i = i$. $Y_i$ values for odd $i$ have positive deviations $r\,i$ from the zero noise case, and $Y_i$ values for even $i$ have negative deviations $-r\,i$ from the zero noise case. The data are illustrated in Fig. 2.2 for the case that $N = 10$ and $r = 0.2$. The data applied suggest the use of a linear model function $y_M(x)$ for modeling the data,

$$y_M = a\,x + b. \tag{2.5}$$

Let us consider now the properties of three error definitions that can be used for the calculation of optimal model parameters $a$ and $b$ in Eq. (2.5).

### 2.2.2 The Chebyshev Error

**Chebyshev Error.** First, we consider the model that results from the use of the Chebyshev error definition for the optimization of the parameters $a$ and $b$ of the linear model Eq. (2.5). For the $(X_i, Y_i)$ data (2.4) considered, the Chebyshev error is given by

$$E_C = \max_{i=1,N} \left| Y_i - y_M(X_i) \right| = \max_{i=1,N} \left| \left(1 - (-1)^i r\right) i - a\, i - b \right|. \tag{2.6}$$

For given values of $r$ and $N$, we can find optimal linear model parameters $a$ and $b$ numerically by determining the $a$ and $b$ values for which $E_C$ becomes minimal. By analyzing the dependence of the optimal $a$ and $b$ values on the parameters $r$ and $N$ of the noise model (2.4), we find $a = 1 - r$ and $b = r\,(N - 1)$ in this way. Thus, the optimal linear model is given by

$$y_M = (1 - r)\, x + r\,(N - 1). \tag{2.7}$$

The minimal value of $E_C$ for the optimal $a$ and $b$ values is given for $N \geq 2$ by

$$E_C = r\,(N - 1). \tag{2.8}$$

**Ease of Use.** Is it difficult to use the Chebyshev criterion for the optimization of model parameters? The variation of the Chebyshev error $E_C$ with the linear model parameter $b$ is shown in Fig. 2.2a, where the optimal value $a = 1 - r = 0.8$ is used for the parameter $a$. This figure is only an example: the corresponding variation of $E_C$ with the parameter $a$ looks similar. In agreement with the conclusion about the optimal $b$ value, $E_C$ attains a minimum at $b = r\,(N - 1) = 1.8$. A disadvantage of using the Chebyshev criterion is that the optimal model parameters $a$ and $b$ can only be found numerically. The reason for that is explained in terms of Fig. 2.2b, which shows that the corresponding derivative $dE_C / db$ jumps at $b = 1.8$. Thus, we cannot analytically determine optimal $a$ and $b$ values because we cannot find the values for which the derivative $dE_C / db = 0$.

**Performance of the Criterion.** The suitability of models obtained by using the Chebyshev criterion for the optimization is illustrated in Fig. 2.2c, which shows Eq. (2.7) for the case considered. This figure shows that the use of the Chebyshev criterion results in a curve such that

$$\begin{aligned} Y_2 - y_M(X_2) = Y_4 - y_M(X_4) = Y_6 - y_M(X_6) = Y_8 - y_M(X_8) \\ = Y_{10} - y_M(X_{10}) = -\left(Y_9 - y_M(X_9)\right). \end{aligned} \tag{2.9}$$

The model (2.7) minimizes the maximum local error, but is does not pay attention to the data trend. In particular, the model behavior disagrees with the data trend for small $i$ values. For $N \to \infty$, Eq. (2.7) implies that $b \to \infty$, which means that the data trend for small $i$ values is not represented at all. Therefore, the use of the Chebyshev criterion is often found to be not the best choice.

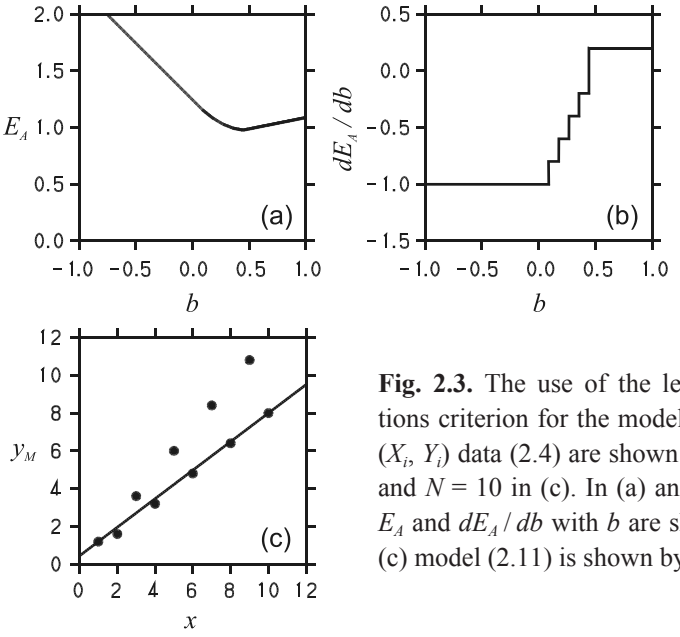



**Fig. 2.3.** The use of the least-absolute-deviations criterion for the model optimization. The ($X_i$, $Y_i$) data (2.4) are shown by *dots* for $r = 0.2$ and $N = 10$ in (c). In (a) and (b), variations of $E_A$ and $dE_A / db$ with $b$ are shown for $a = 0.76$; (c) model (2.11) is shown by a *line*.

### *2.2.3 The Least-Absolute-Deviations Error*

**Least-Absolute-Deviations Error.** Next, let us have a look at the suitability of the least-absolute-deviations error $E_A$ for the optimization of the model (2.5). For the ($X_i$, $Y_i$) data (2.4) considered, $E_A$ is given by

$$E_A = \frac{1}{N}\sum_{i=1}^{N}\left|Y_i - y_M(X_i)\right| = \frac{1}{N}\sum_{i=1}^{N}\left|\left(1-(-1)^i r\right)i - a\,i - b\right|. \tag{2.10}$$

The optimal model parameters $a$ and $b$ that minimize $E_A$ for given $r$ and $N$ values can be found by analyzing the dependence of $E_A$ on $a$ and $b$. By considering the effect of $r$ and $N$ variations one finds the expressions $a = 1 - r\,(N+1)\,/\,(N-1)$ and $b = 2\,N\,r\,(N-1)$. Hence, the optimal linear model is given by

$$y_M = \left(1 - \frac{N+1}{N-1}r\right)x + r\frac{2N}{N-1}. \tag{2.11}$$

The minimal value of $E_A$ for the optimal $a$ and $b$ values is given for $N \geq 2$ by

$$E_A = r\left(\frac{N}{2} - 1\right)\frac{N+1}{N-1}. \tag{2.12}$$

**Ease of Use.** The difficulty of applying the least-absolute-deviations criterion for the optimization of model parameters is addressed in terms of Fig. 2.3. The variation of $E_A$ with the linear model parameter $b$ is shown in Fig. 2.3a, where the optimal value $a = 1 - r\,(N+1)\,/\,(N-1) = 0.76$ is used for the parameter $a$. $E_A$ attains a minimum at $b = 2\,N\,r\,(N-1) = 0.44$. The curve $E_A$ looks like a smooth function, but Fig. 2.3b shows that this view is incorrect: the derivative $dE_A\,/\,db$ is not a smooth function. Consequently, the optimal $a$ and $b$ values cannot be calculated analytically. This fact is definitely a significant disadvantage of the least-absolute-deviations criterion.

**Performance of the Criterion.** The suitability of models obtained by using the least-absolute-deviations criterion for the optimization is illustrated in terms of Fig. 2.3c, which shows the curve (2.11) in comparison with the data considered. This curve (or a closer look at the consequences of Eq. (2.11)) shows that the least-absolute-deviations criterion provides a curve where

$$y_M(X_1) = Y_1, \qquad y_M(X_N) = Y_N. \tag{2.13}$$

For the case considered it turns out that the use of the least-absolute-deviations criterion ends up in a result that can be obtained without using any optimization: any two points can be chosen to define the linear model. There is no reason to assume that the use of the data values at $i = 1$ and $i = N$ is better than other choices. Obviously, such a concept is not helpful in general, which means that the use of the least-absolute-deviations criterion is often found to be not the most convenient choice.

### *2.2.4 The Least-Squares Error*

**Least-Squares Error.** Next, let us consider the properties of the least-squares error $E$. For the $(X_i, Y_i)$ data (2.4) considered, $E^2$ reads

$$E^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - y_M(X_i)\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\left(1-(-1)^i r\right)i - a\,i - b\right]^2. \tag{2.14}$$

The calculation of the optimal model parameters $a$ and $b$ that minimize the error $E^2$ is shown in the beginning of Sect. 2.3.3. According to Eq. (2.61) we obtain $a = 1 - 3\,r\,/\,(N-1)$ and $b = r\,(N+2)\,/\,(N-1)$. Consequently, the optimal model is given by

$$y_M = \left(1 - \frac{3r}{N-1}\right)x + r\frac{N+2}{N-1}. \tag{2.15}$$

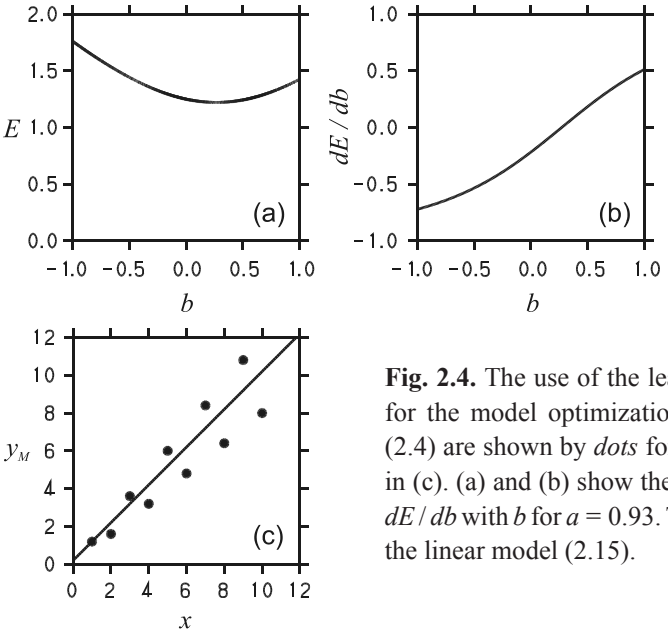



**Fig. 2.4.** The use of the least-squares criterion for the model optimization. The $(X_i, Y_i)$ data (2.4) are shown by *dots* for $r = 0.2$ and $N = 10$ in (c). (a) and (b) show the variations of $E$ and $dE/db$ with $b$ for $a = 0.93$. The *line* in (c) shows the linear model (2.15).

The minimal value of $E^2$ for the optimal $a$ and $b$ values is given for $N \geq 2$ by

$$E^2 = \frac{(N^2-4)r^2}{6}\frac{2N+1}{N-1}. \tag{2.16}$$

**Ease of Use.** Figure 2.4 illustrates the suitability of using the least-squares criterion. The variation of $E$ with the linear model parameter $b$ is shown in Fig. 2.4a, where the optimal model parameter $a = 1 - 3\,r/(N-1) = 0.93$ is used. $E$ attains a minimum at the optimal value $b = r(N+2)/(N-1) = 0.27$. Figure 2.4b shows that the derivative $dE/db$ is a smooth function. Therefore, optimal $a$ and $b$ values can be found analytically (see Sect. 2.3). The latter feature of the least-squares criterion represents a significant advantage of this concept.

**Performance of the Criterion.** Figure 2.4c illustrates the suitability of models obtained by using the least-squares criterion for the optimization. The curve agrees exactly with the first data point,

$$y_M(X_1) = Y_1. \tag{2.17}$$

The model (2.15) does not reveal the disadvantages of the Chebyshev and least-absolute-deviations criteria: it represents the data trend very well, and the model is not simply a curve through any two data points. Therefore, both the good model performance and the property to provide analytical expressions for optimal model parameters suggest the use of the least-squares criterion for the optimization of models.

## 2.3 Optimal Linear Models

Due to shortcomings of alternative concepts that were explained in Sect. 2.2, only the least-squares error concept will be used in the following to determine optimal parameters of models. In this section, we will show how the least-squares error concept can be used for finding optimal linear functions. Before addressing the latter question we introduce in Sect. 2.3.1 useful notation for the description of properties of random data. The use of such definitions of means, variances, and correlations enables an efficient representation of optimal model parameters. We will use these definitions as abbreviations in this chapter. Detailed explanations of the properties of one and several random variables are provided in Chaps. 4 and 10, respectively.

### *2.3.1 Means, Variances, and Correlations*

**Mean Values and Fluctuations.** We consider data values $(X_i, Y_i)$, where $i = 1, N$, this means we have $N$ observations of any two variables $X$ and $Y$. An example for such a data set is given by the data of the noise model (2.4), which are shown in Fig. 2.4. A basic characterization of the variables $X$ and $Y$ is given by the mean values (or expectation values), which are defined by the relations

$$\langle X \rangle = \frac{1}{N}\sum_{i=1}^{N} X_i\,, \qquad \langle Y \rangle = \frac{1}{N}\sum_{i=1}^{N} Y_i\,. \tag{2.18}$$

In general, $X_i$ and $Y_i$ are unequal to their means $<X>$ and $<Y>$, respectively, but there are deviations from the mean values. Such deviations from means (which are called fluctuations) are defined by

$$\widetilde{X}_i = X_i - \langle X \rangle, \qquad \widetilde{Y}_i = Y_i - \langle Y \rangle. \tag{2.19}$$

According to their definition, the mean values of fluctuations disappear,

$$\langle \widetilde{X} \rangle = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \langle X \rangle\right) = \langle X \rangle - \langle X \rangle = 0, \tag{2.20a}$$

$$\langle \widetilde{Y} \rangle = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \langle Y \rangle\right) = \langle Y \rangle - \langle Y \rangle = 0. \tag{2.20b}$$

**Variances.** To understand the relevance of fluctuations we may consider mean values of quadratic variables, which are called second-order moments,

$$\langle X^2 \rangle = \frac{1}{N}\sum_{i=1}^{N} X_i X_i\,, \qquad \langle Y^2 \rangle = \frac{1}{N}\sum_{i=1}^{N} Y_i Y_i\,, \qquad \langle XY \rangle = \frac{1}{N}\sum_{i=1}^{N} X_i Y_i\,. \tag{2.21}$$

The problem related to the consideration of the latter variables is that they do not directly provide the information which we are interested in: information about the amount of fluctuations involved. A more appropriate measure for the intensity of fluctuations is given by the mean values of quadratic deviations from the mean,

$$\left\langle \widetilde{X}\,\widetilde{Y} \right\rangle = \left\langle \left(X - \langle X \rangle\right)\left(Y - \langle Y \rangle\right) \right\rangle = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \langle X \rangle\right)\left(Y_i - \langle Y \rangle\right), \tag{2.22a}$$

$$\left\langle \widetilde{X}^2 \right\rangle = \left\langle \left(X - \langle X \rangle\right)^2 \right\rangle = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \langle X \rangle\right)\left(X_i - \langle X \rangle\right), \tag{2.22b}$$

$$\left\langle \widetilde{Y}^2 \right\rangle = \left\langle \left(Y - \langle Y \rangle\right)^2 \right\rangle = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \langle Y \rangle\right)\left(Y_i - \langle Y \rangle\right). \tag{2.22c}$$

These variables can be written as functions of the moments defined by Eq. (2.21),

$$\left\langle \widetilde{X}^2 \right\rangle = \left\langle X^2 - 2X\langle X \rangle + \langle X \rangle^2 \right\rangle = \left\langle X^2 \right\rangle - \langle X \rangle^2, \tag{2.23a}$$

$$\left\langle \widetilde{Y}^2 \right\rangle = \left\langle Y^2 - 2Y\langle Y \rangle + \langle Y \rangle^2 \right\rangle = \left\langle Y^2 \right\rangle - \langle Y \rangle^2, \tag{2.23b}$$

$$\left\langle \widetilde{X}\widetilde{Y} \right\rangle = \left\langle XY - X\langle Y \rangle - Y\langle X \rangle + \langle X \rangle\langle Y \rangle \right\rangle = \langle XY \rangle - \langle X \rangle\langle Y \rangle. \tag{2.23c}$$

These relations do explicitly demonstrate that the second-order moments involve information about both the amount of fluctuations (given on the left-hand sides), and mean values (given by the last terms). The variables on the left-hand sides are called variances. The variances represent squared variables, i.e., the variances do not directly characterize the range of variations of the random data $X$ and $Y$. The latter information is provided by the standard deviations $<\widetilde{X}^2>^{1/2}$ and $<\widetilde{Y}^2>^{1/2}$ of $X$ and $Y$, respectively.

**Noise Model Example.** An illustration for means and variances defined in this way is given by the expressions that follow from the noise model (2.4),

$$\begin{aligned} &\langle X \rangle = \frac{N+1}{2}, \qquad \langle Y \rangle = \frac{N+1}{2} - \frac{r}{2}, \qquad \left\langle \widetilde{X}\widetilde{Y} \right\rangle = \frac{N^2-1}{12}\left[1 - \frac{3r}{N-1}\right], \\ &\left\langle \widetilde{X}^2 \right\rangle = \frac{N^2-1}{12}, \qquad \left\langle \widetilde{Y}^2 \right\rangle = \frac{N^2-1}{12}\left[1 + 4r^2 - \frac{6r}{N-1} + 3r^2\frac{2N+1}{N^2-1}\right]. \end{aligned} \tag{2.24}$$

These expressions are found by making use of Eqs. (2.18) and (2.22) for the calculation of these quantities. The means $<X>$ and $<Y>$ and the ranges $<X> \pm <\widetilde{X}^2>^{1/2}$ and $<Y> \pm <\widetilde{Y}^2>^{1/2}$ of variations of data values indicated by the corresponding standard deviations are shown in Fig. 2.5. The range of variations $2<\widetilde{X}^2>^{1/2}$ and $2<\widetilde{Y}^2>^{1/2}$ considered in this way does not cover all the data variations. Instead, the standard deviations do only provide the order of magnitude of data variations (see the more detailed explanations in Chap. 4).

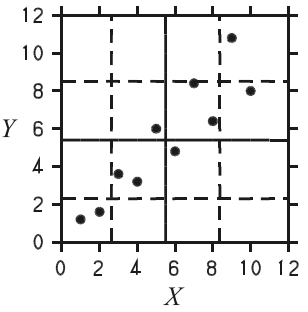



**Fig. 2.5.** An illustration of means and standard deviations of the noise model. The ($X_i$, $Y_i$) data (2.4) are shown by *dots* for $r = 0.2$ and $N = 10$. The *vertical lines* show $\langle X \rangle$ (*solid line*) and $\langle X \rangle \pm \langle \widetilde{X}^2 \rangle^{1/2}$ (*dashed lines*). The *horizontal lines* show $\langle Y \rangle$ (*solid line*) and $\langle Y \rangle \pm \langle \widetilde{Y}^2 \rangle^{1/2}$ (*dashed lines*).

**Correlation Coefficient.** Regarding the consideration of two random numbers $X$ and $Y$, it is relevant to know in which way both variables are related to each other. This information is covered by the cross variance $\langle \widetilde{X}\,\widetilde{Y} \rangle$. The most appropriate way to look at $\langle \widetilde{X}\,\widetilde{Y} \rangle$ is to consider the nondimensional variable

$$r_{XY} = \frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle}{\sqrt{\left\langle \widetilde{X}^2 \right\rangle \left\langle \widetilde{Y}^2 \right\rangle}}, \tag{2.25}$$

which is called the correlation coefficient. This coefficient has the property to be bounded by −1 and 1,

$$-1 \le r_{XY} \le 1. \tag{2.26}$$

To show the validity of this relation we introduce a non-negative function $H(p) \ge 0$ by the definition

$$H(p) = \left\langle \left( \widetilde{X} + p\,\widetilde{Y} \right)^2 \right\rangle = \left\langle \widetilde{X}^2 \right\rangle + 2p \left\langle \widetilde{X}\,\widetilde{Y} \right\rangle + p^2 \left\langle \widetilde{Y}^2 \right\rangle. \tag{2.27}$$

The first two derivatives of $H(p)$ by the parameter $p$ are given by

$$\frac{dH}{dp} = 2\left\langle \widetilde{X}\,\widetilde{Y} \right\rangle + 2p \left\langle \widetilde{Y}^2 \right\rangle, \qquad \frac{d^2H}{dp^2} = 2\left\langle \widetilde{Y}^2 \right\rangle. \tag{2.28}$$

These two derivatives show that $H$ has a minimum value at $p_c = -\langle \widetilde{X}\,\widetilde{Y} \rangle / \langle \widetilde{Y}^2 \rangle$ for which the first-order derivative is equal to zero (provided $\langle \widetilde{Y}^2 \rangle > 0$ as will be assumed here). The minimum $H_{\min}$ of $H(p)$ is given at $p_c$ by

$$H_{\min} = \left\langle \widetilde{X}^2 \right\rangle - \left\langle \widetilde{X}\,\widetilde{Y} \right\rangle^2 \Big/ \left\langle \widetilde{Y}^2 \right\rangle. \tag{2.29}$$

The function $H(p) \ge 0$ for all $p$. Thus, $H(p_c) = H_{\min}$ also has to be non-negative, $H_{\min} \ge 0$. Consequently, we find

$$\frac{\left\langle \widetilde{X}\,\widetilde{Y} \right\rangle^2}{\left\langle \widetilde{X}^2 \right\rangle \left\langle \widetilde{Y}^2 \right\rangle} \le 1, \tag{2.30}$$

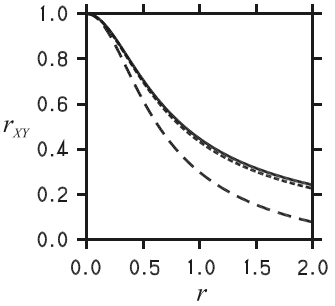



**Fig. 2.6.** The noise model: the dependence of the correlation coefficient $r_{XY}$ on the noise rate $r$. The *solid line* shows the asymptotic curve (2.34). The *dashed lines* show the correlation coefficient in dependence on the sample number $N$ according to Eq. (2.33). *Short dashes* correspond to $N = 100$, *long dashes* to $N = 10$.

which implies the relation (2.26) if the definition (2.25) of $r_{XY}$ is accounted for. The inequality (2.26) represents Schwarz's inequality. Under which conditions do we find specific values of the correlation coefficient? A first specific case is given if $Y_i$ is independent of $X_i$, this means $Y_i$ is not a function of $X_i$ (such variables are called independent variables, see Chap. 4). In this case, $r_{XY}$ is given by

$$r_{XY} = \frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle}{\sqrt{\left\langle \widetilde{X}^2 \right\rangle \left\langle \widetilde{Y}^2 \right\rangle}} = \frac{\left\langle \widetilde{X} \right\rangle \left\langle \widetilde{Y} \right\rangle}{\sqrt{\left\langle \widetilde{X}^2 \right\rangle \left\langle \widetilde{Y}^2 \right\rangle}} = 0. \tag{2.31}$$

A second specific case is given if $Y_i$ is a linear function of $X_i$, $Y_i = c_1 X_i + c_2$, such that $Y_i$ is fully defined by $X_i$. Here, the coefficients $c_1$ and $c_2$ are any constants. In this case, the mean of $Y_i$ is given by $\langle Y_i \rangle = c_1 \langle X_i \rangle + c_2$, and the fluctuation is $\widetilde{Y}_i = c_1 \widetilde{X}_i$. Thus, the correlation coefficient is given by $r_{XY} = c_1 / |c_1|$, which means

$$r_{XY} = \pm 1 \tag{2.32}$$

depending on the sign of $c_1$. It is worth noting that $|r_{XY}| \neq 1$ if $Y_i$ is a function of $X_i$ but not a linear function (see the discussion of Fig. 2.7). Thus, the value of $r_{XY}$ indicates the degree of linear correlation of two random variables.

**Noise Model Example.** Let us consider again the noise model (2.4) to illustrate the variations of $r_{XY}$. By using Eq. (2.24) we find for the correlation coefficient the expression

$$r_{XY} = \frac{1 - \dfrac{3r}{N-1}}{\sqrt{1 + 4r^2 - \dfrac{6r}{N-1} + 3r^2 \dfrac{2N+1}{N^2-1}}}. \tag{2.33}$$

For the case that $N$ is sufficiently large, $r_{XY}$ approaches the limit

$$\lim_{N \to \infty} r_{XY} = \frac{1}{\sqrt{1 + 4r^2}}. \tag{2.34}$$

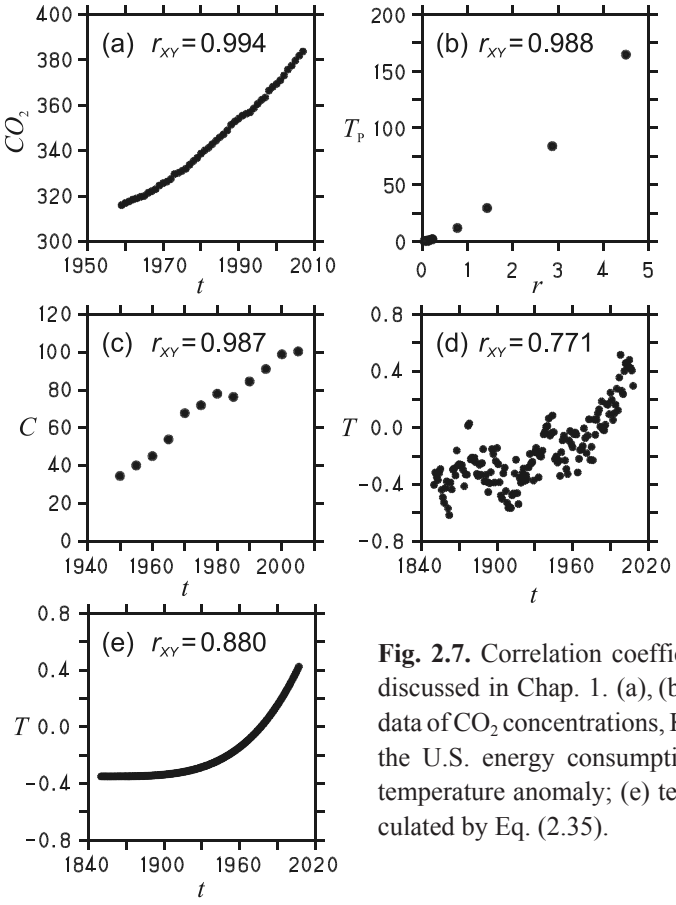



**Fig. 2.7.** Correlation coefficients for problems discussed in Chap. 1. (a), (b), (c), and (d) show data of $CO_2$ concentrations, Kepler's Third Law, the U.S. energy consumption, and the global temperature anomaly; (e) temperature data calculated by Eq. (2.35).

The functions (2.33) and (2.34) are illustrated in Fig. 2.6. The curves shown in this figure support the conclusions of the discussion of the limit cases $r_{XY} = \pm 1$ and $r_{XY} = 0$ in the preceding paragraph. For $r = 0$ (i.e., for $Y_i = X_i$), we have $r_{XY} = 1$. For $r \to \infty$, $Y_i$ becomes independent of $X_i$. Hence, the correlation coefficient $r_{XY} \to 0$ in this case. It is interesting to observe that for the sample numbers $N$ considered the correlation coefficient is close to one for relatively small $r_{XY}$. For example, we find relatively high values $r_{XY} \leq 0.98$ for $r \leq 0.1$, which means for a noise of 10%. The correlation coefficient increases with the number of samples as long as the asymptotic curve (2.34) is not reached.

**Other Examples.** The correlation coefficients of data considered in Chap. 1 are shown in Fig. 2.7. It was shown in Chap. 1 that the $CO_2$ concentration data and the period data related to Kepler's Third Law can be modeled very accurately (the relative error of the corresponding models was below 1%). Thus, these data are hardly affected by randomness, i.e., the minor deviations of these correlation

coefficients from $r_{XY} = 1$ arise, first of all, from the deviations of these data from a linear trend. Regarding the U.S. energy consumption, the deviation of the correlation coefficient from one is implied by the data randomness. It is worth noting that the correlation coefficients in Figs. 2.7a–c are relatively high. Figure 2.7d shows the global temperature anomaly data that are significantly affected by both nonlinearity and randomness. The relevance of nonlinearity and randomness is addressed in terms of Fig. 2.7e. This figure shows the same number of data as Fig. 2.7d, but the temperatures are now calculated from Eq. (1.33),

$$T = -0.35 + \left(\frac{t-1840}{178}\right)^4. \tag{2.35}$$

The correlation coefficient of these non-random temperature data is $r_{XY} = 0.88$. Hence, the main reason for the deviation of the correlation coefficient from one in Fig. 2.7d is the nonlinearity. These examples support the conclusion of the noise model shown in Fig. 2.6: it requires a very high degree of randomness to observe a very small correlation coefficient close to zero.

**Skewness and Flatness.** Means, variances, and the correlation coefficient are the quantities involved in the optimization of linear functions. The optimization of quadratic functions in Sect. 2.4 leads to the need to consider other quantities in addition, like the skewness $m_3$ and the flatness $m_4$ that are defined by

$$m_3 = \left\langle \widetilde{X}^3 \right\rangle \left\langle \widetilde{X}^2 \right\rangle^{-3/2}, \qquad m_4 = \left\langle \widetilde{X}^4 \right\rangle \left\langle \widetilde{X}^2 \right\rangle^{-2}. \tag{2.36}$$

To prepare the use of $m_3$ and $m_4$ in conjunction with the optimization of quadratic functions we derive here a relevant relation between these quantities on the basis of the property $r_{XY}^2 \le 1$ of $r_{XY}$. By replacing in the definition (2.25) of correlation coefficients $\widetilde{X}$ by $\widetilde{X}^2 / < \widetilde{X}^2 > -1$ and $\widetilde{Y}$ by $\widetilde{X} / < \widetilde{X}^2 >^{1/2}$, we find

$$\left\langle \left( \frac{\widetilde{X}^2}{\left\langle \widetilde{X}^2 \right\rangle} - 1 \right) \frac{\widetilde{X}}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}} \right\rangle^2 = \left\langle \frac{\widetilde{X}^3}{\left\langle \widetilde{X}^2 \right\rangle^{3/2}} \right\rangle^2 \le \left\langle \left( \frac{\widetilde{X}^2}{\left\langle \widetilde{X}^2 \right\rangle} - 1 \right)^2 \right\rangle = \left\langle \frac{\widetilde{X}^4}{\left\langle \widetilde{X}^2 \right\rangle^2} \right\rangle - 2 + 1. \tag{2.37}$$

The use of the definitions of $m_3$ and $m_4$ then results in the inequality

$$m_3^2 + 1 \le m_4. \tag{2.38}$$

Under which condition does the equality sign appear here? In correspondence to the conditions for $r_{XY}^2 = 1$, we find this case if $\widetilde{X} / < \widetilde{X}^2 >^{1/2}$ is a linear function of $\widetilde{X}^2 / < \widetilde{X}^2 > -1$, this means

$$\frac{\widetilde{X}}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}} = \frac{1}{m_3} \left( \frac{\widetilde{X}^2}{\left\langle \widetilde{X}^2 \right\rangle} - 1 \right). \tag{2.39}$$

The requirement to use the proportionality factor $1/m_3$ can be seen by multiplying Eq. (2.39) with $\widetilde{X}/<\widetilde{X}^2>^{1/2}$ and taking the average of both sides. The claim that this relation implies the equality sign in Eq. (2.38) can be proven by the multiplication of Eq. (2.39) with $\widetilde{X}^2/<\widetilde{X}^2>$ and averaging: we obtain $m_3 = (m_4 - 1)/m_3$ in this way. A better understanding of the requirement related to Eq. (2.39) can be achieved by multiplying it with $<\widetilde{X}^2>^2$. We find

$$\left\langle \widetilde{X}^3 \right\rangle \widetilde{X} = \left\langle \widetilde{X}^2 \right\rangle \left( \widetilde{X}^2 - \left\langle \widetilde{X}^2 \right\rangle \right), \tag{2.40}$$

where the definition of $m_3$ is applied. This relation represents a quadratic equation for $\widetilde{X}$. By solving this equation, one finds a nonfluctuating value for $\widetilde{X}$. The only possibility that $\widetilde{X}$ does not fluctuate is $\widetilde{X} = 0$, which means the equality sign in Eq. (2.38) can only appear in the absence of $X_i$ fluctuations. Hence, we have

$$0 < m_4 - 1 - m_3^{\,2}, \tag{2.41}$$

if $<\widetilde{X}^2> > 0$, as will be assumed here.

### 2.3.2 Optimal Linear Functions

**Error Formula.** Let us assume that there are $i = 1, N$ data points $(X_i, Y_i)$ that follow, basically, a linear function (for example, the data points given in Fig. 2.4). Our goal is to determine the parameter $a$ and $b$ of the linear model

$$y_M = a\,x + b \tag{2.42}$$

such that this linear function does optimally agree with the data. In particular, we look for $a$ and $b$ parameter values that minimize the least-squares error

$$E^2 = \frac{1}{N}\sum_{i=1}^{N} \left( Y_i - y_M(X_i) \right)^2 = \frac{1}{N}\sum_{i=1}^{N} \left( Y_i - a\,X_i - b \right)^2. \tag{2.43}$$

The error $E^2$ can be rewritten by distributing the sum,

$$\begin{aligned}
E^2 &= \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - 2a\frac{1}{N}\sum_{i=1}^{N} X_i Y_i + a^2 \frac{1}{N}\sum_{i=1}^{N} X_i^{\,2} - 2b\frac{1}{N}\sum_{i=1}^{N} Y_i + 2ab\frac{1}{N}\sum_{i=1}^{N} X_i + b^2 \\
&= \left\langle Y^2 \right\rangle - 2a\left\langle X\,Y \right\rangle + a^2 \left\langle X^2 \right\rangle - 2b\left\langle Y \right\rangle + 2ab\left\langle X \right\rangle + b^2 \\
&= \left\langle \widetilde{Y}^2 \right\rangle + \left\langle Y \right\rangle^2 - 2a\left( \left\langle \widetilde{X}\widetilde{Y} \right\rangle + \left\langle X \right\rangle \left\langle Y \right\rangle \right) + a^2 \left( \left\langle \widetilde{X}^2 \right\rangle + \left\langle X \right\rangle^2 \right) - 2b\left\langle Y \right\rangle + 2ab\left\langle X \right\rangle + b^2 \\
&= \left\langle \widetilde{Y}^2 \right\rangle - 2a\left\langle \widetilde{X}\widetilde{Y} \right\rangle + a^2 \left\langle \widetilde{X}^2 \right\rangle + b^2 - 2b\left\langle Y \right\rangle + \left\langle Y \right\rangle^2 + 2a\left\langle X \right\rangle \left( b - \left\langle Y \right\rangle \right) + a^2 \left\langle X \right\rangle^2 \\
&= \left\langle \widetilde{Y}^2 \right\rangle - 2a\left\langle \widetilde{X}\widetilde{Y} \right\rangle + a^2 \left\langle \widetilde{X}^2 \right\rangle + \left( b - \left\langle Y \right\rangle + a\left\langle X \right\rangle \right)^2.
\end{aligned} \tag{2.44}$$

Here, the second line makes use of the definitions of means (2.18) and second-order moments (2.21). The third line applies Eqs. (2.23), which relate the second-order moments and variances. The fourth line reorganizes the ordering of these terms to prepare the final result that combines the last five terms of the previous line in one quadratic term.

**Critical Points.** The mean values involved here are any numbers that can be calculated from the data points ($X_i$, $Y_i$). Thus, the error $E^2$ is a function of the model parameters $a$ and $b$. A required condition to find a minimum or maximum of $E^2$ is given by the need that the partial derivatives of $E^2$ by $a$ and $b$ disappear (the partial derivative of $E^2$ by $a$ is given by keeping $b$ constant and differentiating $E^2$ by the parameter $a$). These derivatives are given by

$$\frac{\partial E^2}{\partial a} = -2\left\langle \widetilde{X}\widetilde{Y} \right\rangle + 2a\left\langle \widetilde{X}^2 \right\rangle + 2\left\langle X \right\rangle \left( b - \left\langle Y \right\rangle + a\left\langle X \right\rangle \right), \tag{2.45a}$$

$$\frac{\partial E^2}{\partial b} = 2\left( b - \left\langle Y \right\rangle + a\left\langle X \right\rangle \right). \tag{2.45b}$$

The requirement that these partial derivatives have to disappear then results in the following critical points (the last term in (2.45a) disappears because of (2.45b))

$$a = \frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle}, \tag{2.46a}$$

$$b = \left\langle Y \right\rangle - a\left\langle X \right\rangle. \tag{2.46b}$$

The linear model that results from the use of these $a$ and $b$ values is given by

$$y_M = \left\langle Y \right\rangle + \frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle} \left( x - \left\langle X \right\rangle \right). \tag{2.47}$$

Here, the model parameters can be calculated from given data.

**Minimum.** At this point, it is still unclear whether $E^2$ has a minimum for these $a$ and $b$ values, i.e., whether Eq. (2.47) represents an optimal model. This question can be addressed by means of the Second Derivatives Test of Calculus. We define a variable $D$ by the relation

$$D = \frac{\partial^2 E^2}{\partial a^2} \frac{\partial^2 E^2}{\partial b^2} - \left( \frac{\partial^2 E^2}{\partial a \, \partial b} \right)^2. \tag{2.48}$$

According to the Second Derivatives Test, $E^2$ has a minimum at the critical points (2.46) of $a$ and $b$ if

$$\frac{\partial^2 E^2}{\partial a^2} > 0, \qquad \text{and} \qquad D > 0 \tag{2.49}$$

at the critical points. According to Eq. (2.45), the second-order partial derivatives involved in the definition of $D$ are given by

$$\frac{\partial^2 E^2}{\partial a^2} = 2\left\langle \widetilde{X}^2 \right\rangle + 2\left\langle X \right\rangle^2, \qquad \frac{\partial^2 E^2}{\partial b^2} = 2, \qquad \frac{\partial^2 E^2}{\partial a \partial b} = \frac{\partial^2 E^2}{\partial b \partial a} = 2\left\langle X \right\rangle. \quad (2.50)$$

Hence, $\partial^2 E^2 / \partial a^2 > 0$ for all values of the parameter $a$. The use of Eqs. (2.50) in the definition of $D$ shows that

$$D = 4\left\langle \widetilde{X}^2 \right\rangle + 4\left\langle X \right\rangle^2 - 4\left\langle X \right\rangle^2 = 4\left\langle \widetilde{X}^2 \right\rangle, \quad (2.51)$$

which means that $D > 0$ under the condition that $< \widetilde{X}^2 > > 0$, as it will be assumed here. This means that $E^2$ has indeed a local minimum at the critical values (2.46) for $a$ and $b$, which means that Eq. (2.47) represents the optimal model.

**Minimal Error.** The minimal value of the least-squares error $E^2$ can be found by using the critical values for $a$ and $b$ in Eq. (2.44),

$$E^2 = \left\langle \widetilde{Y}^2 \right\rangle - 2\frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle}\left\langle \widetilde{X}\widetilde{Y} \right\rangle + \frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle^2}{\left\langle \widetilde{X}^2 \right\rangle} = \left\langle \widetilde{Y}^2 \right\rangle - \frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle^2}{\left\langle \widetilde{X}^2 \right\rangle}. \quad (2.52)$$

$E^2$ is completely determined by the data properties. Hence, $E^2$ does not say anything about the standard of the linear model applied, which means that $E^2$ is not helpful for the evaluation of the suitability of models. Therefore, no attempt is made to use $E^2$ regarding the applications considered in the following. An evaluation of models obtained can be performed by considering the relative model error, as done in Chap. 1. Rewriting of Eq. (2.52) for the error $E^2$ provides

$$E^2 = \left\langle \widetilde{Y}^2 \right\rangle (1 - r_{XY}{}^2), \quad (2.53)$$

where the definition (2.25) of the correlation coefficient $r_{XY}$ is used. This relation shows that $E^2$ is proportional to $< \widetilde{Y}^2 >$, which is a measure for the typical amount of the $Y$ data randomness. $E^2$ is also proportional to $1 - r_{XY}{}^2$. According to the discussion of properties of the correlation coefficient in Sect. 2.3.1, the limit case $1 - r_{XY}{}^2 = 0$ is given if $Y_i$ is a linear function of $X_i$, and the limit case $1 - r_{XY}{}^2 = 1$ is given if $Y_i$ is independent of $X_i$. Hence, the stronger the dependence of $Y_i$ and $X_i$, the smaller is the error $E^2$.

**Modification.** The approach presented here to derive the optimal linear model (2.47) can be also applied to other linear optimization problems. For example, we may have the problem that there is a relatively complicated function, and we are interested to approximate this function over a certain interval by a linear function. A problem of this kind was considered in Sect. 1.5.3, where we tried to find a linear approximation to the nonlinear temperature-$CO_2$ relation (1.38). Instead of addressing this problem empirically by looking for an appropriate reference point for the Taylor expansion, we may apply the idea of linear optimization described

above to solve such a problem in a much more convenient way. Let $F(x)$ refer to a given complicated function that we would like to approximate by a linear function $F^{(LIN)} = a\,x + b$. Hence, we look for the parameters $a$ and $b$ such that the relation

$$F(x) = a\,x + b \tag{2.54}$$

is optimally satisfied. A first condition for $a$ and $b$ arises from the constraint that Eq. (2.54) has to be satisfied in the integral mean. The integral mean over an interval between $c$ and $d$ is defined for any variable Q (which can be $F$ or $x$) by

$$\langle Q \rangle_I = \frac{1}{d-c} \int_c^d Q(x)\,dx. \tag{2.55}$$

The subscript $I$ of the symbol $<Q>_I$ refers to the integral mean used here instead of the mean (2.18) used above. By integrating Eq. (2.54) between $c$ and $d$ and dividing both sides by $d - c$ we find

$$\langle F \rangle_I = a \langle x \rangle_I + b, \tag{2.56}$$

where we made use of the fact that $<1>_I = 1$. We take the difference between the Eqs. (2.54) and (2.56) to have an equation that does only involve $a$,

$$F(x) - \langle F \rangle_I = a \left( x - \langle x \rangle_I \right). \tag{2.57}$$

The direct integration of this relation does not make sense because we end up with $0 = 0$. Thus, we multiply this condition with $x - <x>_I$ and integrate after that. This leads to the following condition for the parameter $a$,

$$a = \frac{\left\langle \left( F - \langle F \rangle_I \right)\left( x - \langle x \rangle_I \right) \right\rangle_I}{\left\langle \left( x - \langle x \rangle_I \right)^2 \right\rangle_I} = \frac{\langle F\,x \rangle_I - \langle F \rangle_I \langle x \rangle_I}{\langle x^2 \rangle_I - \langle x \rangle_I^{\,2}}. \tag{2.58}$$

The comparison of the optimal parameters $a$ and $b$ obtained by Eqs. (2.58) and (2.56) with the optimal parameters given by Eq. (2.46) reveals the correspondence of the results. The integrals over $x$ involved here can be calculated – we obtain

$$\langle x \rangle_I = \frac{1}{d-c} \int_c^d x\,dx = \frac{1}{2} \frac{d^2 - c^2}{d-c} = \frac{d+c}{2}.$$

$$\left\langle \left( x - \langle x \rangle_I \right)^2 \right\rangle_I = \langle x^2 \rangle_I - \langle x \rangle_I^{\,2} = \frac{1}{3} \frac{d^3 - c^3}{d-c} - \frac{(d+c)^2}{4} = \frac{1}{3} \frac{(d-c)^3 + 3d^2c - 3dc^2}{d-c}$$

$$- \frac{(d+c)^2}{4} = \frac{(d-c)^2}{3} + c\,d - \frac{(d-c)^2}{4} - c\,d = \frac{(d-c)^2}{12}. \tag{2.59}$$

By using these expressions we obtain for $a$ and $b$ in $F^{(LIN)} = a\,x + b$ the relations

$$a = \frac{12}{(d-c)^2} \left[ \langle F\,x \rangle_I - \langle F \rangle_I \frac{d+c}{2} \right], \qquad b = \langle F \rangle_I - a \frac{d+c}{2}. \tag{2.60}$$

### 2.3.3 Examples

Let us illustrate the use of optimal linear models. In particular, we will consider three examples. First, we consider the modeling of the noise model data (2.4), for which the optimal linear model can be directly applied. Second, we will consider Kepler's Third Law data, which requires a data transformation. Third, we will illustrate the application of the optimal linear model (2.54) to continuous functions by the derivation of a linear temperature-$CO_2$ relation (see Sect. 1.5.3).

**Noise Model.** The construction of an optimal linear model for the noise data (2.4) requires the use of Eqs. (2.24) for means and variances in Eqs. (2.46) for $a$ and $b$,

$$a = \frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle} = \frac{12}{N^2-1}\frac{N^2-1}{12}\left[1-\frac{3r}{N-1}\right] = 1-\frac{3r}{N-1}, \tag{2.61a}$$

$$\begin{aligned} b &= \langle Y \rangle - a\langle X \rangle = \frac{N+1}{2} - \frac{r}{2} - \frac{N+1}{2}\left[1-\frac{3r}{N-1}\right] = \frac{r}{2}\frac{3(N+1)-(N-1)}{N-1} \\ &= \frac{r}{2}\frac{2N+4}{N-1} = r\frac{N+2}{N-1}. \end{aligned} \tag{2.61b}$$

Therefore, the optimal linear model is given for this case by

$$y_M = \left(1-\frac{3r}{N-1}\right)x + r\frac{N+2}{N-1}. \tag{2.62}$$

**Kepler's Third Law.** Next, let us use the linear model (2.47) for the derivation of Kepler's Third Law discussed in Sect. 1.2.2. The data values are given in Table 1.2. According to Kepler's Law, the period $T_P$ is related to the mean distance $r$ from the Sun by a power function law,

$$T_P = c\, r^d. \tag{2.63}$$

Here, $c$ and $d$ are model parameters that have to be determined. To enable the use of methods for linear equations, we write Eq. (2.63) as a linear function of the variables $\ln T_P$ and $\ln r$,

$$\ln T_P = d \ln r + \ln c. \tag{2.64}$$

Next, we introduce new variables to enable the direct use of the formulas derived in Sect. 2.3.2, this means write Eq. (2.64) as

$$y_M = a\, x + b, \tag{2.65}$$

where the variables are defined by the relations

$$y_M = \ln T_P, \qquad x = \ln r, \qquad a = d, \qquad b = \ln c. \tag{2.66}$$

The optimal model parameter $a$ and $b$ can be calculated on the basis of Eq. (2.46) and the data from Table 1.2: the $\ln r$ data are the $X_i$ data, and the $\ln T_P$ data are the $Y_i$ data. This calculation provides the optimal values $a = 1.4995$ and $b = 2.8483$. We use Eqs. (2.66) to find for $c$ and $d$ the optimal values

$$c = 17.2584, \qquad d = 1.4995. \tag{2.67}$$

The optimal model for the period $T_P$ is then given by

$$T_P = 17.2584\, r^{1.4995}. \tag{2.68}$$

To compare this result with Eq. (1.8) we use $d = 1.5$ and write

$$T_P = \sqrt{\frac{4\pi^2}{G_S} r^3}, \tag{2.69}$$

where $G_S = 0.1325$ $(10^9$ km$)^3 / a^2$. This result is basically the same as the result $G_S = 0.1324$ $(10^9$ km$)^3 / a^2$ derived in Sect. 1.2.2. Thus, the optimal model (2.69) confirms the result derived in Chap. 1.

**Linear *T-CO*$_2$ Relation.** The third example considered is the derivation of a linear temperature-$CO_2$ relation by means of the optimal linear model for continuous functions. Equation (2.54) reads for this case

$$T = a\, CO_2 + b. \tag{2.70}$$

Here, the temperature $T$ is determined by the function (1.38),

$$T = -0.35 + \left[ \frac{119}{178} + \frac{47}{356} \left( \sqrt{1 + \frac{40}{329}(CO_2 - 316)} - 1 \right) \right]^4. \tag{2.71}$$

The application of Eqs. (2.60) for $a$ and $b$ requires the calculation of the integrals $\langle F \rangle_I$ and $\langle F x \rangle_I$ over the interval considered (i.e., from $c = 316$ to $d = 390$). The convenient way for doing this is to perform the integration numerically. The latter results in

$$\langle F \rangle_I = 0.1637, \qquad \langle F x \rangle_I = 61.6325. \tag{2.72}$$

The use of Eq. (2.60) for the calculation of the parameters $a$ and $b$ in the linear approximation $T^{(LIN)} = a\, CO_2 + b$ for the temperature $T$ leads then to the result

$$T^{(LIN)} = \frac{CO_2 - 333.5}{119.2}. \tag{2.73}$$

This result is very close to the finding $T^{(LIN)} = (CO_2 - 333.3) / 120.6$, which was empirically obtained in Sect. 1.5.3. Figure 2.8 shows an almost perfect agreement between the curves (2.71) and (2.73). The advantage of the approach presented here is that the optimal model parameters are determined for a given interval considered, i.e., there is no need for the adjustment of model parameters.

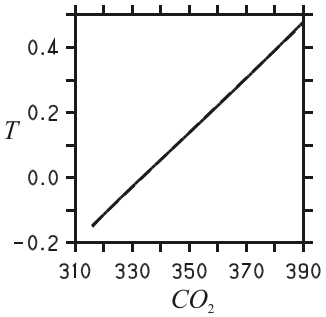



**Fig. 2.8.** The temperature-$CO_2$ relation. The *solid line* shows the exact relation (2.71). The *dashed line* shows the linear approximation (2.73).

## 2.4 Optimal Quadratic Models

Linear models have a limited range of applicability: there are many processes that cannot be described by linear functions. Correspondingly, the extension of the linear methods presented in Sect. 2.3 will be discussed next by addressing the question of how optimal quadratic functions can be constructed. The theory of optimal quadratic functions will be described in Sect. 2.4.1. Two applications will be considered then in Sects. 2.4.2 and 2.4.3.

### *2.4.1 Optimal Quadratic Functions*

**Error Formula.** The development of an optimal quadratic model is more challenging than the development of an optimal linear model. To avoid complicated rewritings, it is the best to consider the following quadratic function,

$$y_M = \left\langle \widetilde{Y}^2 \right\rangle^{1/2} \left( A \frac{\left(x - \langle X \rangle\right)^2}{\left\langle \widetilde{X}^2 \right\rangle} + B \frac{x - \langle X \rangle}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}} + C \right). \tag{2.74}$$

This function can be rewritten into the standard form of a quadratic function,

$$y_M = ax^2 + bx + c, \tag{2.75}$$

where the model parameters $a$, $b$, and $c$ are determined by the relations

$$\begin{aligned} a\left\langle \widetilde{X}^2 \right\rangle &= A\left\langle \widetilde{Y}^2 \right\rangle^{1/2}, \qquad b\left\langle \widetilde{X}^2 \right\rangle = \left\langle \widetilde{Y}^2 \right\rangle^{1/2} \left( B\left\langle \widetilde{X}^2 \right\rangle^{1/2} - 2A\langle X \rangle \right), \\ c\left\langle \widetilde{X}^2 \right\rangle &= \left\langle \widetilde{Y}^2 \right\rangle^{1/2} \left( A\langle X \rangle^2 - B\langle X \rangle \left\langle \widetilde{X}^2 \right\rangle^{1/2} + C\left\langle \widetilde{X}^2 \right\rangle \right). \end{aligned} \tag{2.76}$$

$A$, $B$, and $C$ are unknown parameters that have to be determined. The means are defined by Eq. (2.18) and the variances by Eq. (2.22). The variances $<\widetilde{Y}^2>$ and $<\widetilde{X}^2>$ are assumed to be unequal to zero, which means that there are $Y_i$ and $X_i$ fluctuations. The ratios $(x - <X>) / <\widetilde{X}^2>^{1/2}$ and $y_M / <\widetilde{Y}^2>^{1/2}$ are nondimensional. Hence, the model parameters $A$, $B$, and $C$ are nondimensional, too. The advantage of writing Eq. (2.74) in this way is that the calculation of the model parameters $A$, $B$, and $C$ becomes relatively simple. The least-squares error $E^2$ is given for the quadratic function (2.74) by the expression

$$\begin{aligned} E^2 &= \frac{1}{N}\sum_{i=1}^{N}\left( Y_i - \left\langle \widetilde{Y}^2 \right\rangle^{1/2}\left[ A\frac{\left(X_i - \langle X\rangle\right)^2}{\left\langle \widetilde{X}^2 \right\rangle} + B\frac{X_i - \langle X\rangle}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}} + C \right]\right)^2 \\ &= \frac{\left\langle \widetilde{Y}^2 \right\rangle}{N}\sum_{i=1}^{N}\left( \frac{\widetilde{Y}_i}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}} - A\frac{\widetilde{X}_i^{\,2} - \left\langle \widetilde{X}^2 \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle} - B\frac{\widetilde{X}_i}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}} - \left[ C + A - \frac{\langle Y\rangle}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}} \right]\right)^2 . \end{aligned} \tag{2.77}$$

Here, we used $Y_i = <Y_i> + \widetilde{Y}_i$, and we wrote the second parenthesis term so that this term becomes proportional to $\widetilde{X}_i^{\,2} - <\widetilde{X}^2>$, which vanishes in the mean. By distributing the terms of the sum, the normalized error can be written

$$\begin{aligned} \frac{E^2}{\left\langle \widetilde{Y}^2 \right\rangle} &= \frac{\left\langle \widetilde{Y}^2 \right\rangle}{\left\langle \widetilde{Y}^2 \right\rangle} - 2A\frac{\left\langle \widetilde{Y}\left(\widetilde{X}^2 - \left\langle \widetilde{X}^2 \right\rangle\right)\right\rangle}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}\left\langle \widetilde{X}^2 \right\rangle} + A^2\frac{\left\langle \widetilde{X}^4 - 2\widetilde{X}^2\left\langle \widetilde{X}^2 \right\rangle + \left\langle \widetilde{X}^2 \right\rangle^2 \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle^2} \\ &\quad - 2B\frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}\left\langle \widetilde{X}^2 \right\rangle^{1/2}} + 2AB\frac{\left\langle \widetilde{X}\left(\widetilde{X}^2 - \left\langle \widetilde{X}^2 \right\rangle\right)\right\rangle}{\left\langle \widetilde{X}^2 \right\rangle^{3/2}} + B^2\frac{\left\langle \widetilde{X}^2 \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle} + \left[ C + A - \frac{\langle Y\rangle}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}} \right]^2 . \end{aligned} \tag{2.78}$$

This formula can be written more efficiently as

$$\frac{E^2}{\left\langle \widetilde{Y}^2 \right\rangle} = 1 - 2\,A\,r_{XXY} + A^2\left(m_4 - 1\right) - 2\,B\,r_{XY} + 2\,A\,B\,m_3 + B^2 + \left[ C + A - \frac{\langle Y\rangle}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}} \right]^2 , \tag{2.79}$$

where we applied the abbreviations

$$r_{XY} = \frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}\left\langle \widetilde{Y}^2 \right\rangle^{1/2}},\ r_{XXY} = \frac{\left\langle \widetilde{X}^2\widetilde{Y} \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle\left\langle \widetilde{Y}^2 \right\rangle^{1/2}},\ m_3 = \frac{\left\langle \widetilde{X}^3 \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle^{3/2}},\ m_4 = \frac{\left\langle \widetilde{X}^4 \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle^2}. \tag{2.80}$$

The correlation coefficient $r_{XY}$, skewness $m_3$, and the flatness $m_4$ were already used in Sect. 2.3.1. The coefficient $r_{XXY}$ is defined in correspondence to $r_{XY}$.

**Optimal C.** The optimal $C$ value has to be chosen such that the non-negative bracket term in Eq. (2.79) disappears. Hence, the optimal $C$ value is given by

$$C = \frac{\langle Y \rangle}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}} - A. \tag{2.81}$$

By using this value we find the error to be given by

$$\begin{aligned}
\frac{E^2}{\left\langle \widetilde{Y}^2 \right\rangle} &= 1 - 2Ar_{XXY} + A^2(m_4 - 1) - 2Br_{XY} + 2ABm_3 + B^2 \\
&= (B + Am_3 - r_{XY})^2 - (Am_3 - r_{XY})^2 + 1 - 2Ar_{XXY} + A^2(m_4 - 1) \\
&= (B + Am_3 - r_{XY})^2 + (m_4 - 1 - m_3{}^2)A^2 - 2(r_{XXY} - m_3 r_{XY})A + 1 - r_{XY}{}^2 \\
&= (B + Am_3 - r_{XY})^2 + (m_4 - 1 - m_3{}^2)\left( A - \frac{r_{XXY} - m_3 r_{XY}}{m_4 - 1 - m_3{}^2} \right)^2 \\
&\quad + 1 - r_{XY}{}^2 - \frac{(r_{XXY} - m_3 r_{XY})^2}{m_4 - 1 - m_3{}^2}.
\end{aligned} \tag{2.82}$$

**Optimal *A* and *B*.** Optimal $A$ and $B$ values can be found in correspondence to the calculation of the optimal $C$ value. Equation (2.41) shows that $m_4 - 1 - m_3{}^2 > 0$ if $< \widetilde{X}^2 > > 0$, as is assumed here. Therefore, the first two terms on the right-hand side of Eq. (2.82) are both non-negative. The optimal $A$ and $B$ are the values for which the first two terms are equal to zero, this means the optimal $A$ and $B$ are

$$A = \frac{r_{XXY} - m_3 r_{XY}}{m_4 - 1 - m_3{}^2}, \qquad B = r_{XY} - m_3 A. \tag{2.83}$$

The minimal error that is given for these values reads

$$E^2 = \left\langle \widetilde{Y}^2 \right\rangle \left( 1 - r_{XY}{}^2 - \frac{(r_{XXY} - m_3 r_{XY})^2}{m_4 - 1 - m_3{}^2} \right). \tag{2.84}$$

The comparison with the error (2.53) of the optimal linear model shows that the error of the quadratic model is smaller than the error of the linear model because of the negative last term in Eq. (2.84).

### *2.4.2 Vehicular Stopping Distance*

**Optimal Function.** As a first application of an optimal quadratic function we will consider the vehicular stopping distance problem discussed in Chap. 1. The data values are given in Table 1.3. According to the conclusions of Sect. 1.2.3,

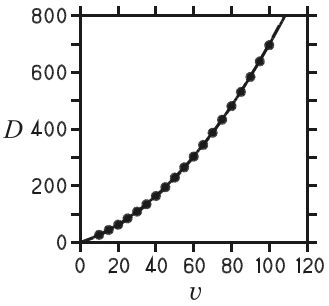



**Fig. 2.9.** The stopping distance $D$ (in feet) as a function of the car's velocity $v$ (in mph). The *solid line* shows the optimal quadratic model (2.86), and the *dashed line* shows the model (2.87), which was derived in Sect. 1.2.3. The data given in Table 1.3 are shown by *dots*.

it is reasonable to assume that the data can be described by a quadratic function

$$D = a v^2 + b v + c. \tag{2.85}$$

The model parameter can be obtained by combining Eqs. (2.76) with the expressions for $A$, $B$, and $C$ derived above. This calculation leads to the result

$$D = 0.0478 v^2 + 2.1771 v + 0.4681 = 0.4681 + \left( 2.1771 + \frac{v}{20.9107} \right) v. \tag{2.86}$$

The last expression represents a rewriting in correspondence to Eq. (1.9) obtained in Sect. 1.2.3,

$$D = \left( 2 + \frac{v}{20} \right) v. \tag{2.87}$$

**Application.** Figure 2.9 shows a comparison between the model (2.86) and the model (2.87) derived in Sect. 1.2.3. This figure does not show any observable difference between both model functions. Nevertheless, a difference is given by the constant 0.4681 that appears in relation (2.86). The latter means that we have at $v = 0$ the stopping distance $D(v = 0) = 0.4681$, which does not make sense. The way to construct an optimal quadratic model cannot account for such conditions. Such a problem can be avoided by developing an optimal linear model for $D / v$ in correspondence to the development in Sect. 1.2.3.

### *2.4.3 $CO_2$ Concentrations*

**Optimal Function.** As a second application of an optimal quadratic model we consider the modeling of the atmospheric $CO_2$ concentration development, which was discussed in Chap. 1. The measured data are given in Table 1.8. As shown in Sect. 1.5.2, the increase of atmospheric $CO_2$ concentrations can be described by a quadratic model

$$CO_2 = a t^2 + b t + c. \tag{2.88}$$

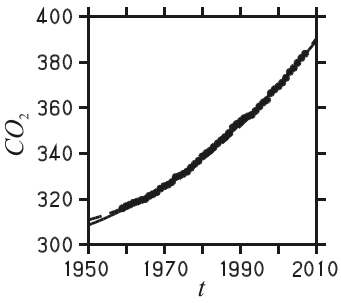



**Fig. 2.10.** Atmospheric $CO_2$ concentration data. The Table 1.8 data are shown as *dots*. The *solid line* shows the optimal quadratic model (2.89); the *dashed line* shows the model (2.90), which was derived in Sect. 1.5.2.

The model parameters can be determined as explained in Sect. 2.4.1. The result of this calculation is given by the function

$$\begin{aligned} CO_2 &= 0.0120\,t^2 - 46.1486\,t + 44690.4788 \\ &= 315.2335 + 0.8446\left(1 + \frac{t-1959}{70.4158}\right)(t-1959). \end{aligned} \tag{2.89}$$

The last line represents a rewriting of the first line in correspondence to Eq. (1.29) obtained in Sect. 1.5.2,

$$CO_2 = 316 + 0.7\left(1 + \frac{t-1959}{47}\right)(t-1959). \tag{2.90}$$

**Application.** Figure 2.10 presents a comparison of the optimal model (2.89) with the empirical model (2.90), which was derived in Sect. 1.5.2. The performance of both models is very similar. It seems that the use of the empirical model (2.90) is more appropriate due to the very good agreement with the data values close to $t = 1959$. However, a closer look shows that the optimal quadratic model (2.89) provides a better overall prediction, in particular for data between $t = 1980$ and $t = 1990$. Thus, the optimal model (2.89) actually is the slightly better model.

## 2.5 Optimal Power and Exponential Models

Linear and quadratic functions are applicable to many modeling problems, but there is also a variety of problems that cannot be handled on this basis. Therefore, we consider here the construction of optimal power and exponential functions. In some cases it is possible to rewrite such functions such that optimization methods for linear functions can be applied (see, e.g., the optimization of redefined model variables in Kepler's Third Law discussed in Sect. 2.3.3). This approach may become inapplicable if we have the (relatively usual) case that we have to account

for a reference point $(x_0, y_0)$ as given in the functions

$$y_M = y_0 + a(x - x_0)^b. \tag{2.91a}$$

$$y_M = y_0 + a(e^{b(x-x_0)} - 1). \tag{2.91b}$$

Here, $a$ and $b$ are any model parameters. It may be seen that $y_M = y_0$ at $x = x_0$. It will be assumed that $x_0$ is any known reference point – usually it is not difficult to find an appropriate value. For a known $y_0$ value, we can write the power function model (2.91a) as a linear function of redefined variables, but this approach does not work regarding the exponential function model (2.91b). The modeling of the noisy data of the global temperature anomaly discussed in Sect. 1.5.2 showed that it may be impossible to find an appropriate value for $y_0$ for the power function model (2.91a), such that this model cannot be linearized. To handle such cases we need an approach that allows the direct optimization of Eqs. (2.91). The way to handle this question will be described in the following, this means we will find optimal $y_0$, $a$, and $b$ parameter values for Eqs. (2.91). The problems considered here represent examples for the way to address such questions: there are many similar problems that can be solved correspondingly. The optimization of Eqs. (2.91) will be explained and illustrated in Sects. 2.5.1 and 2.5.2. An application to real data, the modeling of the global temperature increase, will be discussed in Sect. 2.5.3.

### *2.5.1 Optimal Power Functions*

**Error Definition.** We assume that there are $i = 1, N$ data points $(X_i, Y_i)$ that suggest the use of the power function model (2.91a). The least-squares error reads for the model considered

$$\begin{aligned} E^2 &= \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - y_M(X_i)\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - y_0 - a(X_i - x_0)^b\right)^2 \\ &= {y_0}^2 - 2y_0\left\langle Y - a(X - x_0)^b\right\rangle + \left\langle\left(Y - a(X - x_0)^b\right)^2\right\rangle \\ &= \left(y_0 - \langle Y\rangle + a\langle U\rangle\right)^2 - \left(\langle Y\rangle - a\langle U\rangle\right)^2 + \left\langle(Y - aU)^2\right\rangle. \end{aligned} \tag{2.92}$$

Here, we used in the second line the definitions (2.18) of means. In the third line we used the abbreviation $U = (X - x_0)^b$. It is assumed that $x_0$ is given such that $x_0 \neq X_i$ for all $i = 1, N$. This condition avoids problems with the existence of $U$ for negative $b$ values. If $x_0$ has to be equal to one $X_i$ value, then we may disregard this data point regarding the calculation of model parameters. In the following, the goal will be to find optimal $y_0$, $a$, and $b$ values that minimize $E^2$.

**Optimal $y_0$.** The optimal $y_0$ value is the value for which the non-negative first term on the right-hand side of Eq. (2.92) is equal to zero, this mean

$$y_0 = \langle Y \rangle - a\langle U \rangle. \tag{2.93}$$

The error can be written for this case

$$\begin{aligned} E^2 &= \left\langle (Y - aU)^2 \right\rangle - \left( \langle Y \rangle - a\langle U \rangle \right)^2 \\ &= \left\langle Y^2 \right\rangle - \langle Y \rangle^2 - 2a\left( \langle UY \rangle - \langle U \rangle \langle Y \rangle \right) + a^2 \left( \langle U^2 \rangle - \langle U \rangle^2 \right). \end{aligned} \tag{2.94}$$

The use of the variance definitions (2.23) enables a simpler writing of $E^2$,

$$\begin{aligned} E^2 &= \left\langle \tilde{Y}^2 \right\rangle - 2a\left\langle \tilde{U}\tilde{Y} \right\rangle + a^2 \left\langle \tilde{U}^2 \right\rangle = \left\langle \tilde{Y}^2 \right\rangle + \left\langle \tilde{U}^2 \right\rangle \left( a^2 - 2a \frac{\left\langle \tilde{U}\tilde{Y} \right\rangle}{\left\langle \tilde{U}^2 \right\rangle} \right) \\ &= \left\langle \tilde{Y}^2 \right\rangle + \left\langle \tilde{U}^2 \right\rangle \left( a - \frac{\left\langle \tilde{U}\tilde{Y} \right\rangle}{\left\langle \tilde{U}^2 \right\rangle} \right)^2 - \frac{\left\langle \tilde{U}\tilde{Y} \right\rangle^2}{\left\langle \tilde{U}^2 \right\rangle}. \end{aligned} \tag{2.95}$$

**Optimal $a$ and $b$.** The optimal $a$ value is the value for which the non-negative second term on the right-hand side of Eq. (2.95) is equal to zero. Thus, $a$ becomes

$$a = \frac{\left\langle \tilde{U}\tilde{Y} \right\rangle}{\left\langle \tilde{U}^2 \right\rangle}. \tag{2.96}$$

The error $E^2$ can be written then

$$E^2 = \left\langle \tilde{Y}^2 \right\rangle - \frac{\left\langle \tilde{U}\tilde{Y} \right\rangle^2}{\left\langle \tilde{U}^2 \right\rangle} = \left\langle \tilde{Y}^2 \right\rangle \left( 1 - \frac{\left\langle \tilde{U}\tilde{Y} \right\rangle^2}{\left\langle \tilde{U}^2 \right\rangle \left\langle \tilde{Y}^2 \right\rangle} \right) = \left\langle \tilde{Y}^2 \right\rangle (1 - r_{UY}{}^2). \tag{2.97}$$

Here, $r_{UY}$ refers to the correlation coefficient of $U$ and $Y$. The optimal parameters $y_0$, $a$, and the error $E^2$ obtained here correspond to the results for a linear function: see Eqs. (2.46) and (2.53). This is not surprising because the variables $U$, $a$, and $y_0$ used here correspond to $X$, $a$, and $b$ in the linear function discussed in Sect. 2.3.2. The difference to the optimization of a linear function is that the error $E^2$ given by Eq. (2.97) depends on the model parameter $b$. Unfortunately, there is no way to find the optimal $b$ value analytically. Therefore, this value has to be determined numerically as the value for which $E^2$ becomes minimal. The optimal parameters $a$ and $y_0$ can be obtained then on the basis of Eqs. (2.96) and (2.93).

**Table 2.1** Data used for the illustration of the optimization of a power function.

| $X$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 10.5 | 3.6 | 2.4 | 1.8 | 1.5 | 1.4 | 1.3 | 1.2 | 1.17 | 1.15 |

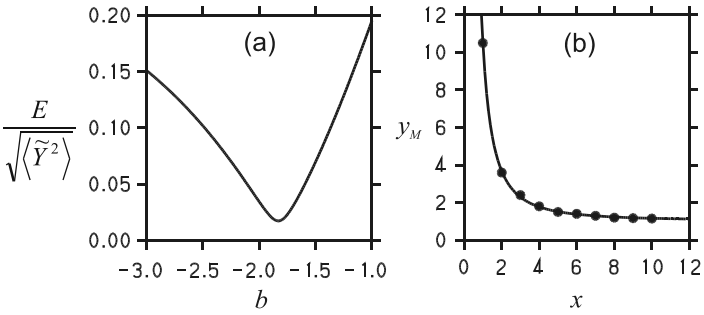



**Fig. 2.11.** An optimal power function model. (a) The error $E/\langle \tilde{Y}^2 \rangle^{1/2}$ given by Eq. (2.97); this function has a minimum at $b = -1.8312$; (b) the *dots* show the $(X_i, Y_i)$ data values given in Table 2.1, the *line* shows the corresponding optimal power function model (2.98).

**Application.** Let us illustrate the application of this approach. The data shown in Table 2.1 represent a data set that can be described by a power function. The reference point $x_0$ has to be unequal to the data points applied. For simplicity, we use $x_0 = 0$. The error $E/\langle \tilde{Y}^2 \rangle^{1/2}$ determined by Eq. (2.97) is shown in Fig. 2.11a as a function of $b$. The minimum of this error is given at $b = -1.8312$. For this value of $b$ we find via Eqs. (2.96) and (2.93) the optimal values $a = 9.4681$ and $y_0 = 1.0237$. The resulting power function model is then given by

$$y_M = 1.0237 + \frac{9.4681}{x^{1.8312}}. \tag{2.98}$$

Fig. 2.11 shows that this model represents the data considered very well.

### 2.5.2 Optimal Exponential Functions

**Error Definition.** As a second example, let us address the optimization of the exponential function (2.91b). For this function, the least-squares error reads

$$\begin{aligned} E^2 &= \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - y_M(X_i)\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - y_0 - a\left(e^{b(X_i - x_0)} - 1\right)\right)^2 \\ &= {y_0}^2 - 2\,y_0\left\langle Y - a\left(e^{b(X-x_0)} - 1\right)\right\rangle + \left\langle \left(Y - a\left(e^{b(X-x_0)} - 1\right)\right)^2 \right\rangle \\ &= \left(y_0 - \langle Y\rangle + a\langle U\rangle\right)^2 - \left(\langle Y\rangle - a\langle U\rangle\right)^2 + \left\langle (Y - aU)^2 \right\rangle. \end{aligned} \tag{2.99}$$

Here, we used the abbreviation $U = \exp[b\,(X - x_0)] - 1$. This error formula is equal to Eq. (2.92). Thus, the optimal values of $y_0$, $a$, and $b$, and the resulting expression for $E^2$ are equal to the formulas given in Sect. 2.5.1.

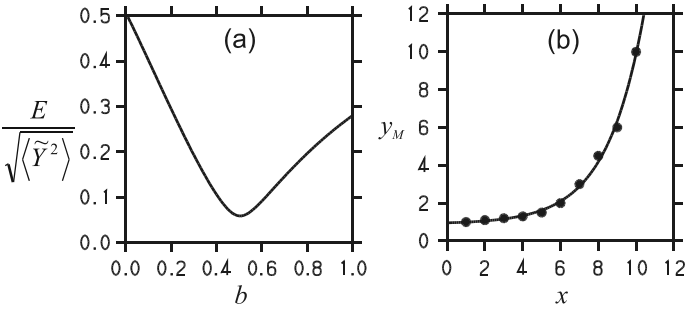



**Fig. 2.12.** An optimal exponential function model. (a) The error $E/<\tilde{Y}^2>^{1/2}$ given by Eq. (2.97); this function has a minimum at $b = 0.5043$; (b) the *dots* show the data given in Table 2.2, the *line* shows the optimal exponential function model (2.100).

**Application.** The following example illustrates the use of this approach. The data shown in Table 2.2 and Fig. 2.12 suggest the use of an exponential function for the modeling of these data. We may use the first point for the definition of $x_0 = 1$. The error $E/<\tilde{Y}^2>^{1/2}$ determined by Eq. (2.97) is shown in Fig. 2.12a. The minimum of this function is given at $b = 0.5043$. The optimal values of $a$ and $y_0$ are then given by $a = 0.0963$ and $y_0 = 0.9959$. With these model parameters we find the exponential model

$$y_M = 0.9959 + 0.0963(e^{0.5043(x-1)} - 1). \tag{2.100}$$

The comparison of this function with the data in Fig. 2.12 shows that the model (2.100) accurately represents the data considered.

### *2.5.3 Global Warming*

**Global Temperature Modeling.** Let us use now the power function modeling approach from Sect. 2.5.1 for finding an optimal model for real data: we will consider again the global temperature increase discussed in Chap. 1. The temperature data are given in Table 1.7. The discussion in Sect. 1.5.2 showed that the development of the global temperature anomaly $T$ in time $t$ may be well described

**Table 2.2** Data used for the illustration of the optimization of an exponential function.

| $X$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 1.0 | 1.1 | 1.2 | 1.3 | 1.5 | 2.0 | 3.0 | 4.5 | 6.0 | 10.0 |

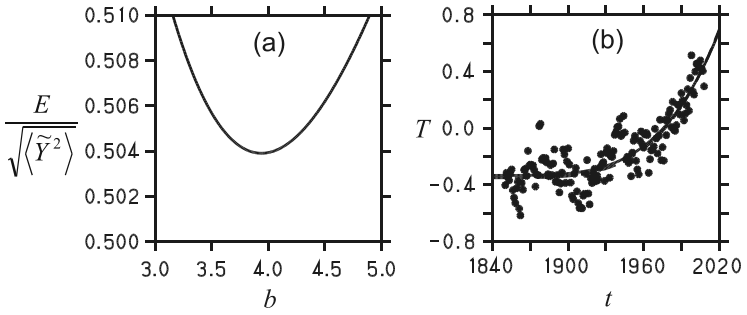



**Fig. 2.13.** The global temperature increase. (a) The error $E / <\widetilde{Y}^2>^{1/2}$ given by (2.97); this function has a minimum at $b = 3.9402$; (b) the *dots* show the data from Table 1.7, the *solid line* shows the optimal power function model (2.102). The *dashed line* shows the model (2.103) derived in Sect. 1.5.2.

by the power function

$$T = T_0 + a(t - t_0)^b = T_0 + \left(\frac{t - t_0}{c}\right)^b . \tag{2.101}$$

The last expression is written according to the result from Eq. (1.33) in Sect. 1.5.3. The parameter $c$ introduced here is related to $a$ and $b$ by $c = 1 / a^{1/b}$.

**Optimal Model.** The methodology for optimizing a power function described in Sect. 2.5.1 can be directly applied to this case: $T$, $T_0$, $t$, and $t_0$ correspond to $y$, $y_0$, $x$, and $x_0$, respectively. We set $t_0 = 1840$ in accordance with the conclusions of Sect. 1.5.2. The error $E / <\widetilde{Y}^2>^{1/2}$ is shown in Fig. 2.13a as a function of $b$. The minimum of this error is given at $b = 3.9402$. Instead of the model parameter $a$ we consider here $c = 1 / a^{1/b}$, for which we find the optimal value $c = 179.1856$. The value $T_0$ is found to be $T_0 = -0.3339$. These data result in the power function

$$T = -0.3339 + \left(\frac{t - 1840}{179.1856}\right)^{3.9402} . \tag{2.102}$$

**Application.** The performance of the optimal model (2.102) is shown in Fig. 2.13 in comparison to the data values. The model

$$T = -0.35 + \left(\frac{t - 1840}{178}\right)^4 \tag{2.103}$$

derived in Sect. 1.5.2 is also shown here. There is only a very minor difference between both models for the interval $t = 1840$ to $t = 1870$. Due to the scatter of data values, it is impossible to conclude that one of the two models is better than the other one: the performance of both models is good. The significant advantage of the optimal model (2.102) is that there is no information required in addition to the data values. On the other hand, the development of the model (2.103) obtained

before is not comparably simple. The modeling approach applied in Sect. 1.5.2 works for this case, but it may not work for other cases.

## 2.6 Summary

This chapter addressed the question of how it is possible to find a model that does optimally agree with observations involving randomness. Let us summarize the conclusions regarding the questions considered in Sect. 2.1, this means the questions about the most convenient optimization concept, the difficulty of using this concept, and the benefits of the development of optimal models.

**Optimization Concept.** There are many concepts available for minimizing the deviations between a model and data – so which concept should we apply? The optimization of linear relations represents the most important application area of optimization concepts. Therefore, various optimization concepts were compared for such a case (the linear noise model). Both the least-absolute-deviations error and Chebyshev error do not provide analytical expressions for the optimal model parameters. The Chebyshev error does not correctly describe the data trend, and the least-absolute-deviations error provides a curve that passes two data points (this means we end up with a result that can be obtained without using any optimization). The least-squares error does not reveal such shortcomings. In addition, the use of this concept provides analytical expressions for the optimal parameters. Thus, only the least-squares error concept was used in the following. It is worth noting that this concept does not provide a tool for the evaluation of the performance of models (we obtain an optimal model that may perform well or not). This question has to be addressed by calculating the relative error of model predictions (see Chap. 1).

**Ease of Use.** How difficult is the use of the optimization concept? The point is that there would be not much interest in applying an optimization concept if its use turns out to be very complicated (we would consider alternatives). The application of the least-squares error concept for the optimization of linear functions (of redefined variables) is relatively simple. This concept can be also used for finding analytically optimal parameters for quadratic functions. The error concept can be also applied to the development of optimal power and exponential functions, although we have to numerically determine now the minimum of the error (the error is a function of one variable). Therefore, the error concept can be applied to all the basic functions considered in Fig. 1.17. The optimization concept can be also applied to more complex functions by determining optimal model parameter values numerically. However, such calculations are not simple anymore, in particular if we have to deal with three or more variables.

**Benefits.** The significant advantage of the least-squares optimization concept is that this concept usually provides appropriate model parameters in a very efficient way. There is no need to perform expensive studies of the effects of parameter variations, and there is no need for the consideration of averaged values, which were used in Chap. 1 to model the global temperature anomaly. The parameters obtained are at least as good as the results of specific investigations of effects of parameter variations. However, this optimization concept does not apply any other condition than the requirement to minimize the deviations between a model and observations. Thus, it may be the case that we obtain a nonzero vehicular stopping distance at a zero velocity. This concept does also not provide integer exponents that we prefer, for example regarding the presentation of Kepler's Third Law and the time dependence of the global temperature anomaly. Such questions can be addressed at best by adjusting the results for optimal model parameters according to the needs (by fixing integer exponents of power functions) and optimizing the remaining model parameters.

## 2.7 Exercises

**2.2.1** Consider the data sets (d) and (f) given in Table 1.4. Assume that the data can be modeled in terms of the function $y_M = a\,x$.

a) Use the least-squares error, the Chebyshev error, and the least-absolute-deviations error to find the optimal value of $a$ for the two data sets.

b) Comment on the suitability of the three error concepts for the modeling of both data sets.

**2.2.2** Consider the data set (f) in Table 1.4. Assume that the data can be modeled in terms of the function $y_M = a\,x$.

a) Suggest two reasonable global error concepts in addition to the errors considered in exercise 2.2.1.

b) Calculate the optimal value of $a$ on the basis of your two global error concepts and the least-squares error.

c) Comment on the suitability of your two error concepts.

**2.2.3** Consider the energy consumption data in Table 1.1. Assume that the data can be modeled in terms of the function $C = a\,t + b$.

a) Use the data point at 1950 to express $b$ as a function of $a$.

b) Calculate the optimal parameter $a$ on the basis of the least-squares error, the Chebyshev error, and the least-absolute-deviations error.

c) Graph the three curves. Comment on the suitability of error concepts.

**2.3.1** Consider the least-squares error $E^2$ for a linear function $y_M = a\,x + b$ (see Eq. (2.44) in Sect. 2.3.2),

$$E^2 = \left\langle \widetilde{Y}^2 \right\rangle - 2a\left\langle \widetilde{X}\widetilde{Y} \right\rangle + a^2\left\langle \widetilde{X}^2 \right\rangle + \left(b - \langle Y \rangle + a\langle X \rangle\right)^2.$$

a) Determine directly the optimal $b$ value that minimizes $E^2$. Do not apply the Second Derivatives Test.

b) Determine in the same way the optimal $a$ value that minimizes $E^2$.

**2.3.2** Consider the quadratic model (1.29) for the atmospheric $CO_2$ concentration development derived in Sect. 1.5.2,

$$CO_2 = 316 + 0.7\left(1 + \frac{t - 1959}{47}\right)(t - 1959).$$

a) Use the approach for the linearization of continuous functions described at the end of Sect. 2.3.2 for the development of a linear model for $CO_2$ for the period 1990–2010.

b) Demonstrate the suitability of the linear model obtained by graphing both the quadratic model and the linear model.

**2.3.3** Consider the U.S. population data from 1790 to 1890 given in Table 1.5.

a) Calculate the correlation coefficient $r_{XY}$ for these ($P$, $t$) data.

b) Consider the same $t$ data. Replace the $P$ data by the function values $P_M = 0.0039 \exp[(t - 1790)/35]$. Find $r_{XY}$ for the ($P_M$, $t$) data.

c) Use the data ($P$, $t$) from Table 1.5 to generate new ($\ln P$, $t$) data. Calculate the correlation coefficient $r_{XY}$ for the ($\ln P$, $t$) data.

d) Replace in the latter data $\ln P$ by $\ln P_M$. Find $r_{XY}$ for the ($\ln P_M$, $t$) data.

e) What do these four correlation coefficients tell us?

**2.3.4** Assume that this data set can be modeled by a power function $y_M = a\,x^b$.

| $X$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $Y$ | 0.4 | 0.65 | 0.86 | 1.06 | 1.23 |

a) Introduce new variables that are linearly related. Find the optimal model parameter values for this linear model by using the least-squares error.

b) Graph the resulting optimal power function $y_M = a\,x^b$ and the data.

**2.3.5** Assume that this data set can be modeled by a power function $P = P_0 + b\,t^a$.

| $t$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $P$ | 100 | 101 | 108 | 140 | 230 |

a) Introduce new variables that are linearly related. Find the optimal model parameter values for this linear model by using the least-squares error.

b) Graph the resulting optimal power function $P = P_0 + b\,t^a$ and the data.

**2.3.6** Consider again the energy consumption data in Table 1.1. Assume that the data can be modeled in terms of the function $C = a\,t + b$.
a) Use the least squares error concept to find an optimal model.
b) Graph the optimal model, the model $C = 91 + 1.25\,(t - 1995)$ obtained in Sect. 1.2.1, and the data.
c) Comment on the performance of the optimal model in comparison to the model $C = 91 + 1.25\,(t - 1995)$.

**2.3.7** Consider again the U.S. population data from 1790 to 1890, which are given in Table 1.5. Assume that the data can be described by the exponential function $P = c\,\exp(dt)$.
a) Write the model as a linear model formulated in terms of new variables.
b) Use the least squares error concept to find an optimal model.
c) Graph the population data, the optimal model obtained in b), and the model $P = 0.0039\,\exp[(t - 1790)/35]$ obtained in Sect. 1.4.1.
d) Comment on the performance of the optimal model in comparison to the model $P = 0.0039\,\exp[(t - 1790)/35]$.

**2.3.8** Consider again the U.S. population data from 1790 to 1890, which are given in Table 1.5. Assume that the data can be described by the logistic function $P = A / [1 + \exp(C\,(t - B))]$.
a) Write the model as a linear model formulated in terms of new variables. Assume in consistency with the results of Sect. 1.4.1 that $A = 0.12$.
b) Use the least squares error concept to find an optimal model.
c) Graph the population data, the optimal model obtained in b), and the logistic model $P = 0.12 / [1 + \exp(-\,(t - 1890)/29)]$ obtained in Sect. 1.4.1.
d) Comment on the performance of the optimal model in comparison to the model $P = 0.12 / [1 + \exp(-\,(t - 1890)/29)]$.

**2.3.9** Consider again the global temperature anomaly data given in Table 1.7 for every ten years (1850, 1860, …, 2000). Assume that these temperature data can be described by the function $T = A + [(t - 1840)/B]^4$.
a) Use the least-squares error concept to find an optimal model.
b) Graph the resulting optimal model and the data.

**2.4.1** Consider the optimal quadratic model described in Sect. 2.4.1.
a) What are the requirements for a data set considered such that the optimal quadratic model reduces to a linear optimal model?
b) Show that the linear optimal model obtained by the simplification of the optimal quadratic model is equal to the linear optimal model derived in Sect. 2.3.2.

**2.4.2** Consider the optimal quadratic model described in Sect. 2.4.1.
a) Develop a numerical scheme for using the formulas in Sect. 2.4.1 to find the parameters that minimize the least-squares error of this model.
b) Calculate the optimal model parameters for the data set (0, 4), (1, 8.4), (2, 16.2), and (3, 26.3).

**2.4.3** Consider the total stopping distance data in exercise 1.3.1.
a) Find the optimal quadratic model for this data set.
b) Comment on the suitability of this optimal quadratic model.

**2.5.1** Consider the power function model $y_M = y_0 + a\,x^b$. Here, the exponent $b$ is known, and the coefficients $y_0$ and $a$ have to be optimized.
a) Determine the values of the parameters $y_0$ and $a$ that minimize the least-squares error of this model.
b) Assume that $b = 2$. Calculate the optimal model for the data set (0, 0), (1, 3.4), (2, 14.1), and (3, 31.3).

**2.5.2** Consider the model $y_M = y_0 + \ln(x / x_0)^a$. Here, $y_0$ and $a$ are parameters that have to be optimized. Assume that $x_0$ is known.
a) Determine the values of the parameters $y_0$ and $a$ that minimize the least-squares error of this model.
b) Calculate the optimal values of $y_0$ and $a$ for the data set (1, 0), (2, 3.1), (3, 4.9), and (4, 6.2). Assume that $x_0 = 1$.

**2.5.3** Consider the model $y_M = y_0 + a\,x^b\,e^{-x}$. The parameters $y_0$, $a$, and $b$ have to be optimized.
a) Reduce the problem such that the least-squares error does only depend on one variable. Hint: follow the approach presented in section 2.5.1.
b) Explain the use of these formulas for the calculation of $y_0$, $a$, and $b$.

**2.5.4** Both the exponential function model $y_M = y_0 + a\,(\exp[b\,(x - x_0)] - 1)$ and the power function model $y_M = y_0 + a\,(x - x_0)^b$ require the setting of $x_0$. The parameters $y_0$, $a$, and $b$ can be optimized as explained in Sect. 2.5.
a) Can we have the same optimal exponential function model for two different choices of $x_0$? Show the parameter relations if the answer is yes.
b) Can we have the same optimal power function model for two different choices of $x_0$? Show the parameter relations if the answer is yes.

**2.5.5** Consider the exponential function model $y_M = y_0 + a\,(\exp[b\,(x - x_0)] - 1)$. The parameters $y_0$, $a$, and $b$ have to be optimized.
a) Develop a numerical scheme for using the formulas in Sect. 2.5.2 to find the $y_0$, $a$, and $b$ values that minimize the least-squares error of this model.
b) Calculate the optimal $y_0$, $a$, and $b$ for the data set (1, 1.95), (2, 1.91), (3, 1.87), and (4, 1.84). Use $x_0 = 0$.

**2.5.6** Consider the power function model $y_M = y_0 + a\,(x - x_0)^b$. The parameters $y_0$, $a$, and $b$ have to be optimized.

a) Develop a numerical scheme for using the formulas in Sect. 2.5.1 to find the parameters $y_0$, $a$, and $b$ that minimize the least-squares error of this model.

b) Calculate the optimal $y_0$, $a$, and $b$ for the data set (1, 9), (2, 2), (3, 1.3), and (4, 1.1). Use $x_0 = 0$.

**2.5.7** Consider again the exercise 2.5.6.

a) Redo the calculation for $x_0 = 0.5$.

b) Redo the calculation for $x_0 = 0.8$.

c) Calculate the maximum relative error for the power function models that apply $x_0 = 0$, $x_0 = 0.5$, and $x_0 = 0.8$. Which of the three $x_0$ considered is the best choice?

**2.5.8** Continue with the exercises 2.5.6 and 2.5.7.

a) Present at least two ideas of how it is possible to design a model that performs better than the three models considered in exercises 2.5.7.

b) Demonstrate the suitability of your approach by designing a model that performs better than the three models considered in exercises 2.5.7.

# 3 Deterministic States

The deterministic methods discussed in Chap. 1 were presented for problems that require the modeling of observations of one variable (like the total stopping distance of cars, the energy consumption, the atmospheric $CO_2$ concentration, or the global temperature anomaly). However, most real problems are characterized by relations that involve several variables. Such relations have to provide the correct dimension of variables considered (for example, a characteristic time scale of any problem has to be calculated by a relation that provides a time). This constraint implies a reduction of all possible relations between variables, i.e., it reduces the original complexity of problems significantly (for example, it may require that a certain variable cannot be involved in a relation). The technique that provides such a problem reduction will be described in this chapter. This approach is extremely helpful, but it cannot completely solve the problem. The equations obtained in this way still involve unknown parameters that have to be determined by means of observations. Hence, the methods to be described in this chapter usually represent the first step before using the modeling approaches presented in Chap. 1 to determine unknown parameters.

Section 3.1 illustrates the need for the development of a method for the analysis of problems that involve several variables. Section 3.2 introduces the theoretical basis for the reduction of several-variable problems, i.e., dimensional analysis. The relevant concept of similarity is also explained. This concept provides the theoretical basis for one of the most important applications of dimensional analysis: the possibility to study the properties of real systems (e.g., the forces on an aircraft) by means of much more efficient investigations of little model systems. Applications of these techniques to the solution of problems of low complexity, medium complexity, and relatively high complexity will be described in Sects. 3.3, 3.4, and 3.5, respectively. Section 3.6 summarizes the basic features, advantages, and limitations of the approach presented.

S. Heinz, *Mathematical Modeling*, DOI 10.1007/978-3-642-20311-4_3,

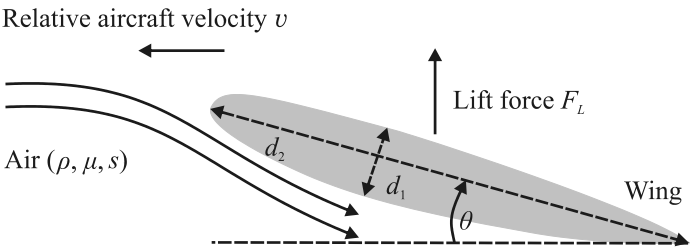



**Fig. 3.1.** An illustration of factors that affect the lift force of aircrafts.

## 3.1 Motivation

**Lift.** Let us consider aircraft flight as an example to illustrate the need to deal with problems that involve many variables. In particular, let us try to find a formula for the dependence of the lift force $F_L$ on various factors. The lift force $F_L$ is a force that is generated to overcome the aircraft weight, i.e., $F_L$ is the force that makes the aircraft fly. Figure 3.1 shows an illustration of how the lift force is generated and affected by various factors. The airflow passing under the wing is deflected downward by the bottom surface of the wing. The downward deflected air is pushing against the surrounding air, the surrounding air is stopping the air deflected downward and pushing this air upward (this process is similar to the flight of a balloon: if air is rushing out of a balloon, the reaction is that the balloon is forced away). The more air is deflected downward, the more lift will be created. The theoretical explanation for this process is provided by Newton's Third Law – for every action there is an equal and opposite reaction (Newton's Laws will be presented in Chap. 7).

**Lift Influence Factors.** There are several factors that may affect the magnitude of the lift force $F_L$: see the illustration in Fig. 3.1. The lift force will depend on the wing geometry. The wing shape may be characterized by two length scales $d_1$ and $d_2$ (which is called the chord length). The lift force will also depend on the relative aircraft velocity $v$ and the angle of attack $\theta$. The relative aircraft velocity $v$ is the speed of an aircraft relative to the airmass in which it flies, i.e., $v$ is the magnitude of the vector difference between the aircraft velocity and the air velocity. The angle of attack $\theta$ refers to the angle of inclination of the wing. The lift force will also depend on the properties of the fluid considered: the air density $\rho$, the air viscosity $\mu$, and the speed of sound $s$. There is one additional factor that may affect the lift force: gravity – which acts in the opposite direction of the lift force. The parameter that accounts for the gravity force is the gravity acceleration $g$. By extending the analysis presented below by the inclusion of the parameter $g$ one

can prove that gravity may contribute to $F_L$ via the ratios $(d_2 g)^{1/2}/v$ or $(d_1 g)^{1/2}/v$. To assess the relevance of these ratios let us consider $d_2 \leq 10$ m (the value of $d_1$ is smaller than $d_2$). The latter assumption results in the inequality $(d_2 g)^{1/2} \leq 20$ km/h. Most airplanes fly at much higher velocities. Thus, the ratio $(d_2 g)^{1/2}/v$ will be negligibly small such that we can disregard the effect of $g$. Correspondingly, we expect $F_L$ as a function of seven variables,

$$F_L = F_L(v, d_1, d_2, \theta, \rho, \mu, s). \tag{3.1}$$

**Experimental Studies.** Let us consider the use of measurements for the evaluation of the lift force in dependence of these seven variables. One way to address this problem is to measure the dependence of $F_L$ on one parameter, where the other six parameters are kept constant. To have an idea about the number of measurements required, let us consider the case that we need five data points (experiments) to characterize the influence of one parameter on $F_L$. To study the dependence of $F_L$ on a second variable, we need $5^2$ measurements. The evaluation of the dependence of $F_L$ on 7 parameters then requires $5^7 = 78{,}125$ experiments! Obviously, the costs of such a number of experiments will be huge. Apart from that, five measurements are not always sufficient to accurately determine data trends, and there will be also a significant influence of randomness in such data.

**Problem Considered.** The discussion in the preceding paragraph showed that experiments cannot be used directly in general to determine the dependence of variables like $F_L$ on a relatively large number of factors. A first essential step is needed: we have to reduce the number of influence factors as much as possible. Experiments, theoretical results, or simulation results can be used then to study the detailed properties of reduced equations.

## 3.2 Dimensional Analysis and Similarity

A way to address the problems described in the previous section is given by the application of dimensional analysis, i.e., the analysis of implications of the most basic property of variables to have a dimension. Let us see how this works.

### *3.2.1 Buckingham's Theorem*

**Dimension.** We are interested in variables that can be observed, i.e., we are interested in measurable variables. The most basic property of such variables is that measurable variables have a dimension. For example, the dimension of the lift

**Table 3.1** Dimensions of physical variables in the LMT system.

| Variable | Dimension | Variable | Dimension |
|---|---|---|---|
| Mass | $M$ | Work | $M\,L^2\,T^{-2}$ |
| Length | $L$ | Pressure | $M\,L^{-1}\,T^{-2}$ |
| Time | $T$ | Power | $M\,L^2\,T^{-3}$ |
| Frequency | $T^{-1}$ | Angle | $M^0\,L^0\,T^0$ |
| Velocity | $L\,T^{-1}$ | Velocity of sound | $L\,T^{-1}$ |
| Acceleration | $L\,T^{-2}$ | Density | $M\,L^{-3}$ |
| Force | $M\,L\,T^{-2}$ | Dynamic viscosity | $M\,L^{-1}\,T^{-1}$ |
| Energy | $M\,L^2\,T^{-2}$ | Kinematic viscosity | $L^2\,T^{-1}$ |

force $F_L$ considered above is mass times acceleration, this means mass times length divided by a squared time. Table 3.1 shows examples for the dimension of other variables. To work with dimensions of variables we need a mathematical formulation of dimensions. Dimensions are represented by a certain unit system. One unit system is given by the *LMT* system. Here, $L$ refers to any unit length $L$, $M$ refers to any unit mass, and $T$ refers to any unit time. Examples are given by $L = 1$ m, $M = 1$ g, and $T = 1$ s. Another example is the choice $L = 1$ km, $M = 1$ kg, and $T = 10^3$ s. The definition of such units is a requirement for measurements. To measure any variable actually means to compare the variable directly or indirectly with an appropriate standard (with a certain unit of the measurement). Here, $L$, $M$, and $T$ serve as units such that every measurement of a length, mass, or time can be represented as a multiple of the corresponding unit. Dimensions of variables are then given as certain combinations of units. For example, the dimension of the lift force $F_L$ considered above is given by $M\,L\,T^{-2}$. The *LMT* system can be used to study a wide range of problems in mechanics. Nevertheless, the *LMT* system is not general in the sense that it can be used for all problems. Additional standards have to be taken into account if problems are studied that involve, for example, heat transfer or electromagnetic systems. However, we will consider only the *LMT* system in the following. This sort of analysis is adequate to explain all the basic features of dimensional analysis. Additional properties can be easily taken into account if required.

**Dimensionally Correct Equation.** The most basic constraint that we can use for the calculation of variables is that the formula applied must have the correct dimension of the variable considered: formulas have to be dimensionally correct. Let us illustrate the benefits of this constraint regarding the calculation of the lift force $F_L = F_L(v, d_1, d_2, \theta, \rho, \mu, s)$ as a function of all these parameters. We know that $F_L$ has the dimension of mass times acceleration, i.e.

$$F_L \sim \frac{M\,L}{T^2}. \tag{3.2}$$

To relate the right-hand side to variables involved in this problem we rewrite Eq. (3.2) in the following way:

$$F_L \sim \frac{M}{L^3}\frac{L^2}{T^2}L^2. \tag{3.3}$$

The first term $M/L^3$ is written in terms of the dimension of the air density $\rho$. This term accounts for the influence of $M$ on $F_L$. The second term $L^2/T^2$ is written in terms of the dimension of $v^2$. This terms accounts for the influence of $T$ on $F_L$. The third term $L^2$ is written in terms of the dimension of $d_2{}^2$ (another possible choice would be the use of $d_1$ instead of $d_2$). This term accounts for the effect of $L$ on $F_L$. According to Eq. (3.3), we may expect the following formula for the lift force $F_{L,}$

$$F_L = \frac{C_L}{2}\rho v^2 d_2{}^2. \tag{3.4}$$

Here, $C_L$ is the so-called lift coefficient. The inclusion of the factor 1/2 in Eq. (3.4) corresponds to the general practice of writing this formula. The variable $p = \rho v^2/2$ is called the dynamic pressure. Thus, Eq. (3.4) can also be written $F_L = C_L\, p\, d_2{}^2$. Equation (3.4) includes only a few variables of the list of all possible variables that may affect $F_L$. Does this mean that the other variables are irrelevant? The latter conclusion is incorrect because the other variables may affect $F_L$ via $C_L$. However, $C_L$ is a nondimensional number, this means $C_L$ is independent of $M$, $T$, and $L$. A closer look at the other variables shows that $C_L$ can be, for example, a function like

$$C_L = C_L\left(\theta, \frac{d_1}{d_2}, \frac{v}{s}, *\right). \tag{3.5}$$

The star indicates that $C_L$ may depend on additional nondimensional parameters. An important conclusion at this point is that a dimensionally correct equation for $F_L$ can be written as a relation between nondimensional products,

$$\frac{F_L}{\rho v^2 d_2{}^2} = \frac{1}{2}C_L\left(\theta, \frac{d_1}{d_2}, \frac{v}{s}, *\right), \tag{3.6}$$

where Eq. (3.4) was applied. In other words, only equations that can be written in terms of nondimensional products are dimensionally correct. Equation (3.6) can be written in a more general way. Let us introduce the following abbreviations for the nondimensional products involved in Eq. (3.6),

$$P_1 = \frac{F_L}{\rho v^2 d_2{}^2}, \qquad P_2 = \theta, \qquad P_3 = \frac{d_1}{d_2}, \qquad P_4 = \frac{v}{s}. \tag{3.7}$$

By using these abbreviations, Eq. (3.6) can be written

$$P_1 = f(P_2, P_3, P_4, \cdots). \tag{3.8}$$

**Buckingham's Theorem.** Equation (3.8) does not only apply to the lift force problem considered here. It can be used for any nondimensional products, and it may involve many additional nondimensional products like $P_5$, $P_6$, ⋯. Equation (3.8) is called Buckingham's Theorem (Buckingham 1914). This theorem states the following. Every dimensionally correct equation can be written like Eq. (3.8): a dimensionally correct equation represents a relation between nondimensional products. A dimensionally correct equation is also called a dimensionally homogeneous equation. Such equations are correct regardless of the system of units in which the variables are measured. The nondimensional products involved in Eq. (3.8) are considered to be independent, this means no nondimensional product can be represented by a combination of other nondimensional products. The nondimensional products involved are considered to represent a complete set of nondimensional products. This assumption means that the nondimensional products are independent, and that all possible nondimensional products can be obtained through combinations of these products. A good way to see whether a particular set of nondimensional products is complete or not is to ask whether all the relevant variables are involved, and whether every product involves one physical variable that does not appear in any other nondimensional product.

**Open Question.** The relevance of Buckingham's theorem is that we cannot work with any equations, but only some equations (which are in consistency with Buckingham's theorem) are correct. The reduction of the number of equations that can be used implies a significant simplification of problems. However, Buckingham's Theorem does not explain how we can derive the most general nondimensional equation for any specific problem: this theorem does only tell us that we have to relate nondimensional products. Thus, there is the question of how it is possible to determine all nondimensional products that should be considered in a nondimensional equation for a given problem.

### *3.2.2 Dimensional Analysis*

**Dimensional Analysis.** Let us address the latter question regarding the lift force problem (the same approach can be applied to other problems: see below). We consider the following general product involving variables that may affect $F_L$,

$$F_L{}^a v^b d_1{}^c d_2{}^d \theta^e \rho^f \mu^g s^h. \tag{3.9}$$

Here, the exponents $a$, $b$, $c$, $d$, $e$, $f$, $g$, and $h$ are any unknown (positive, negative, or zero) real numbers. The condition for having a nondimensional product is then

$$F_L^{\,a}\, v^b\, d_1^{\,c}\, d_2^{\,d}\, \theta^e\, \rho^f\, \mu^g\, s^h = c_1. \tag{3.10}$$

The symbol $c_1$ on the right-hand side refers to any constant that is independent of $L$, $M$, and $T$. To understand the implications of Eq. (3.10) we have to consider the dimensions of the variables involved. By replacing the variables by their corresponding dimensions given in Table 3.1, Eq. (3.10) can be written

$$\left(\frac{M\,L}{T^2}\right)^a \left(\frac{L}{T}\right)^b (L)^c\,(L)^d \left(L^0\,M^0\,T^0\right)^e \left(\frac{M}{L^3}\right)^f \left(\frac{M}{LT}\right)^g \left(\frac{L}{T}\right)^h = c_2. \tag{3.11}$$

Here, $c_2$ is a constant that is independent of $L$, $M$, and $T$ and different from $c_1$. The angle $\theta$ is independent of $L$, $M$, and $T$. Therefore, $\theta$ is replaced by $L^0\,M^0\,T^0$ here. Equation (3.11) can be written in a more convenient way by combining terms that involve $L$, $M$, and $T$,

$$L^{a+b+c+d-3f-g+h}\, M^{a+f+g}\, T^{-2a-b-g-h} = c_2. \tag{3.12}$$

This condition is satisfied if the exponents of $L$, $M$, and $T$ disappear, this means if

$$0 = a + b + c + d - 3f - g + h, \tag{3.13a}$$

$$0 = a + f + g, \tag{3.13b}$$

$$0 = -2a - b - g - h. \tag{3.13c}$$

Equations (3.13) can be solved in different ways. Here, we use these equations for the calculation of $b$, $d$, and $f$, this means $b$, $d$, and $f$ are considered as dependent variables, and $a$, $c$, $e$, $g$, and $h$ represent independent variables. Equations (3.13b) and (3.13c) can be used for the calculation of $f$ and $b$ as functions of independent variables, but Eq. (3.13a) also involves the dependent variables $f$ and $b$. Thus, we rewrite Eq. (3.13a) by means of the expressions for $f$ and $b$ that are provided by Eqs. (3.13b) and (3.13c),

$$0 = a - \left(2a + g + h\right) + c + d + 3\left(a + g\right) - g + h = 2a + c + d + g. \tag{3.14}$$

Correspondingly, $b$, $d$, and $f$ can be calculated by the equations

$$d = -2a - c - g, \tag{3.15a}$$

$$f = -a - g, \tag{3.15b}$$

$$b = -2a - g - h. \tag{3.15c}$$

The use of these expressions in Eq. (3.10) then provides

$$F_L^{\,a}\, v^{-2a-g-h}\, d_1^{\,c}\, d_2^{\,-2a-c-g}\, \theta^e\, \rho^{-a-g}\, \mu^g\, s^h = c_1. \tag{3.16}$$

The latter relation can be also written

$$\left(\frac{F_L}{\rho v^2 d_2{}^2}\right)^a \left(\frac{d_1}{d_2}\right)^c \theta^e \left(\frac{\mu}{\rho d_2 v}\right)^g \left(\frac{s}{v}\right)^h = c_1. \tag{3.17}$$

**Buckingham's Theorem.** This relation is the general condition for having a nondimensional product. It identifies five numbers

$$\frac{F_L}{\rho v^2 d_2{}^2}, \qquad \theta, \qquad \frac{d_1}{d_2}, \qquad \frac{\mu}{\rho d_2 v}, \qquad \frac{s}{v}. \tag{3.18}$$

All these numbers have to be independent of $L$, $M$, and $T$ to satisfy Eq. (3.17) for all $a$, $c$, $e$, $g$, and $h$. Hence, we found five nondimensional products. The latter fact can be easily proven by using the dimensions of these variables in each of these products. The nondimensional products obtained are independent because each product contains a variable ($F_L$, $\theta$, $d_1$, $\mu$, $s$, respectively) that does not appear in any other product. The nondimensional products (3.18) represent a complete list of nondimensional products because all variables considered are involved, and the products (3.18) are independent. Therefore, Eq. (3.18) provides all the nondimensional products that we need to find the general nondimensional equation for the lift force $F_L$. According to Buckingham's Theorem, this nondimensional equation is given by

$$f\left(\frac{F_L}{\rho v^2 d_2{}^2}, \theta, \frac{d_1}{d_2}, \frac{s}{v}, \frac{\mu}{\rho d_2 v}\right) = 0. \tag{3.19}$$

By assuming that we can solve for $F_L/(\rho v^2 d_2{}^2)$, Eq. (3.19) can be written

$$F_L = \frac{C_L}{2} \rho v^2 d_2{}^2, \tag{3.20}$$

where the lift coefficient $C_L$ is given by

$$C_L = C_L\left(\theta, \frac{d_1}{d_2}, \frac{v}{s}, \frac{\mu}{\rho d_2 v}\right). \tag{3.21}$$

This result is the most general expression for $C_L$. Equation (3.5) is generalized by this expression for $C_L$ by the consideration of $\mu/(\rho d_2 v)$. The derivation of the structure of the lift force formula in this way represents a significant advantage compared to the consideration of the lift force $F_L = F_L(v, d_1, d_2, \theta, \rho, \mu, s)$ as a function of seven variables.

**First Variation.** Expression (3.20) for $F_L$ combined with the lift coefficient (3.21) is not the only result that can be obtained by the application of dimensional analysis. In order to illustrate the possible variations, let us consider $b$, $c$, and $f$ as dependent variables instead of $b$, $d$, and $f$ used before. According to Eq. (3.15), the

equations for $b$, $c$, and $f$ are given by

$$c = -2a - d - g, \tag{3.22a}$$

$$f = -a - g, \tag{3.22b}$$

$$b = -2a - g - h. \tag{3.22c}$$

The use of the latter relations in Eq. (3.10) provides

$$F_L{}^a v^{-2a-g-h} d_1{}^{-2a-d-g} d_2{}^d \theta^e \rho^{-a-g} \mu^g s^h = \\ = \left(\frac{F_L}{\rho v^2 d_1{}^2}\right)^a \left(\frac{d_2}{d_1}\right)^d \theta^e \left(\frac{\mu}{\rho d_1 v}\right)^g \left(\frac{s}{v}\right)^h = c_1. \tag{3.23}$$

For this case, the corresponding list of nondimensional products is given by

$$\frac{F_L}{\rho v^2 d_1{}^2}, \qquad \theta, \qquad \frac{d_2}{d_1}, \qquad \frac{\mu}{\rho d_1 v}, \qquad \frac{s}{v}. \tag{3.24}$$

Instead of Eq. (3.20) we obtain then

$$F_L = \frac{C_L}{2} \rho v^2 d_1{}^2, \tag{3.25}$$

where the lift coefficient $C_L$ is given by

$$C_L = C_L\left(\theta, \frac{d_2}{d_1}, \frac{v}{s}, \frac{\mu}{\rho d_1 v}\right). \tag{3.26}$$

**Second Variation.** Let us consider a second variation where $a$, $b$, and $f$ are the dependent parameters. Equations (3.15) show that $a$, $b$, and $f$ are determined by

$$a = -\frac{1}{2}(c + d + g), \tag{3.27a}$$

$$f = -a - g = \frac{1}{2}(c + d - g), \tag{3.27b}$$

$$b = -2a - g - h = c + d - h. \tag{3.27c}$$

Here, Eq. (3.27a) was used to write the other two relations as functions of independent variables. The use of these relations in Eq. (3.10) then provides

$$F_L{}^{-(c+d+g)/2} v^{c+d-h} d_1{}^c d_2{}^d \theta^e \rho^{(c+d-g)/2} \mu^g s^h = \\ = \left(\frac{\rho v^2 d_1{}^2}{F_L}\right)^{c/2} \left(\frac{\rho v^2 d_2{}^2}{F_L}\right)^{d/2} \theta^e \left(\frac{\mu^2}{\rho F_L}\right)^{g/2} \left(\frac{s}{v}\right)^h = c_1. \tag{3.28}$$

The corresponding list of nondimensional products is given for this case by

$$\frac{F_L}{\rho v^2 d_1{}^2}, \qquad \frac{F_L}{\rho v^2 d_2{}^2}, \qquad \theta, \qquad \frac{\rho F_L}{\mu^2}, \qquad \frac{s}{v}. \tag{3.29}$$

This list of nondimensional products cannot be used to represent $F_L$ as a function of other parameters because $F_L$ appears in three of the products in Eq. (3.29).

**Observations.** Let us summarize the observations regarding the variations of analysis results described in the preceding three paragraphs.

- A closer look at the influence of different choices of independent and dependent variables shows that the choice of an independent variable implies that this variable appears in only one nondimensional product, whereas the choice of a dependent variable implies that this variable can be found in many nondimensional products. For example, for our last case considered we used $c$, $d$, $e$, $g$, $h$ as independent variables and $a$, $b$, and $f$ as dependent variables. The variables $d_1$, $d_2$, $\theta$, $\mu$, $s$ related to $c$, $d$, $e$, $g$, $h$ appear only in one nondimensional product, whereas the variables $F_L$, $v$, $\rho$ related to $a$, $b$, and $f$ are found in several nondimensional products.
- There are many different ways to write the results of nondimensional analysis depending on the choice of dependent variables: new nondimensional products can be obtained by using different dependent variables. These differences do not pose any problem as long as the exponent related to the variable $F_L$ (which we like to calculate) is taken as an independent variable. Otherwise, $F_L$ is found in many nondimensional products.
- A procedure for obtaining new products in a simple way is to consider powers of given nondimensional products, or to multiply or divide nondimensional products with (powers of) other nondimensional products; e.g., the operations

$$\frac{F_L}{\rho v^2 {d_2}^2}\left(\frac{d_2}{d_1}\right)^2, \qquad \theta, \qquad \left(\frac{d_1}{d_2}\right)\left(\frac{d_2}{d_1}\right)^2, \qquad \frac{\mu}{\rho d_2 v}\frac{d_2}{d_1}, \qquad \frac{s}{v}, \tag{3.30}$$

enable the transition from the list (3.18) to the list (3.24.)

### *3.2.3 Similarity*

**The Problem.** A usual problem, which will be illustrated again regarding the lift force problem, is the following one. After applying dimensional analysis, we found for $F_L$ the formula

$$F_L = \frac{C_L}{2}\rho v^2 {d_2}^2, \tag{3.31}$$

where the lift coefficient $C_L$ is given by the function

$$C_L = C_L\left(\theta, \frac{d_1}{d_2}, \frac{v}{s}, \frac{\nu}{d_2 v}\right). \tag{3.32}$$

Here, the kinematic viscosity $\nu = \mu / \rho$ is used to simplify the writing. The lift coefficient $C_L$ is an unknown function: we do only know that $C_L$ depends on four parameters. What we need to know is the value of $C_L$ for any specific conditions [e.g., $C_L(10, 0.1, 0.4, 10^{-9})$, see Sect. 3.5], to be able to calculate the lift force $F_L$ for this case. Let us call this value

$$C_L{}^R = C_L{}^R\left(\theta^R, \frac{d_1{}^R}{d_2{}^R}, \frac{v^R}{s^R}, \frac{\nu^R}{d_2{}^R v^R}\right). \tag{3.33}$$

Here, the superscript $R$ refers to the conditions (the values of variables) of the real system. Usually, investigations of $C_L$ for the real system (the aircraft) are way too complicated and expensive. Thus, we try to use a little model system to determine $C_L$. In the little system we measure, e.g., $C_L(10, 0.1, 0.004, 10^{-6})$, or, more general

$$C_L{}^M = C_L{}^M\left(\theta^M, \frac{d_1{}^M}{d_2{}^M}, \frac{v^M}{s^M}, \frac{\nu^M}{d_2{}^R v^M}\right). \tag{3.34}$$

Here, the superscript $M$ refers to the conditions (e.g., the length scales $d_1{}^M$ and $d_2{}^M$) of the model system. Do we calculate the lift force for the real system correctly if we use such a measured $C_L{}^M$ in the lift force formula (3.31)? This approach will provide an incorrect result, in general, because of the following. We need the function $C_L$ at the $R$-values [e.g., $C_L(10, 0.1, 0.4, 10^{-9})$], but we provide $C_L$ at the $M$-values [e.g., $C_L(10, 0.1, 0.004, 10^{-6})$], which will be different from the $R$-values, in general. This means, we consider $C_L$ at the wrong function values.

**The Solution.** The way to overcome the latter problem is to measure $C_L$ in the model system at the values required for the real system, this means we have to make sure that all $M$-ratios are equal to the corresponding $R$-ratios,

$$\theta^M = \theta^R, \qquad \frac{d_1{}^M}{d_2{}^M} = \frac{d_1{}^R}{d_2{}^R}, \qquad \frac{v^M}{s^M} = \frac{v^R}{s^R}, \qquad \frac{\nu^M}{d_2{}^M v^M} = \frac{\nu^R}{d_2{}^R v^R}. \tag{3.35}$$

For this case of equal arguments of $C_L$ we have

$$C_L{}^M = C_L{}^R, \tag{3.36}$$

which means we have the $C_L$ value that is required in the lift force formula (3.31). The last two relations show that all nondimensional products involved have to be equal in the model and in the real system. Such systems are called similar systems. It is worth noting that the similarity of systems does not only require the similarity of geometries (equal $\theta$ and $d_1 / d_2$) – it requires the equality of all nondimensional products involved in the problem considered. Equations (3.35) represent the design conditions for the development of the model system. The design of the model has to be performed such that $\theta^M = \theta^R$ and $(d_1 / d_2)^M = (d_1 / d_2)^R$ are satisfied.

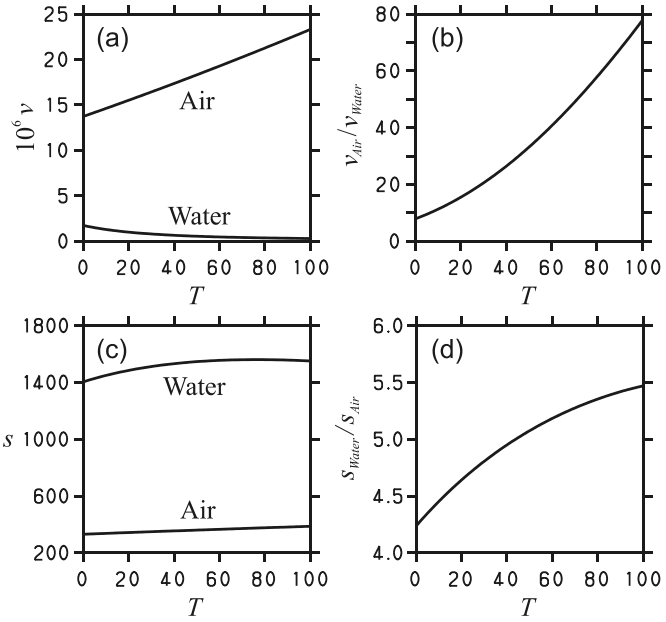



**Fig. 3.2.** Air and water properties. (a) and (c) show the kinematic viscosity $\nu$ and speed of sound $s$ of air and water in dependence on the temperature $T$ (in ºC); (b) and (d) show the ratios $\nu_{Air}/\nu_{Water}$ and $s_{Water}/s_{Air}$, respectively.

The latter is not very complicated, in general. It means that the model has to look like the real system. Written as conditions for the velocity ratio $v^M/v^R$, the last two conditions of Eq. (3.35) that have to be satisfied are given by

$$\frac{v^M}{v^R}=\frac{s^M}{s^R}, \qquad \frac{v^M}{v^R}=\frac{\nu^M}{\nu^R}\frac{{d_2}^R}{{d_2}^M}. \tag{3.37}$$

In essence, it is not simple at all to satisfy these two conditions. By replacing the velocity ratio in the second relation according to the first relation, these conditions can be also written as

$$\frac{v^M}{v^R}=\frac{s^M}{s^R}, \qquad \frac{{d_2}^M}{{d_2}^R}=\frac{\nu^M}{\nu^R}\frac{s^R}{s^M}. \tag{3.38}$$

The second condition shows that it needs another fluid to obtain a model that is smaller than the real system.

**Kinematic Viscosity and Speed of Sound of Air and Water.** Let us consider the kinematic viscosity $\nu$ and the speed of sound $s$ of air and water to prepare the use of these properties for the discussion of the realization of design conditions.

The viscosity of air at standard atmospheric pressure is given (in $m^2/s$) by

$$\nu_{Air} = 3.4484\,10^{-6}\left(-0.0683\alpha^3 + 2.0712\alpha^2 + 2.9786\alpha - 1\right). \tag{3.39}$$

Here, $\alpha = (T + 273.15) / 273.15$, and $T$ refers to the temperature in Celsius. This formula represents a polynomial curve fit to a data set that is given in the SFPE Handbook of Fire Protection Engineering (1995). The range of validity of this formula is 100°K ≤ $T$ ≤ 1600°K. Due to thermodynamics (Heinz 2003), the speed of sound (in m/s) in dry air at standard atmospheric pressure is

$$s_{Air} = 331.5\sqrt{\alpha}\,. \tag{3.40}$$

This formula can be used at least for –40ºC ≤ $T$ ≤ 100ºC. According to Chenlo et al. (2004), the viscosity of water is given (in $m^2/s$) by

$$\nu_{Water} = 0.09607\,10^{-6} e^{2.9/\alpha^3}\,. \tag{3.41}$$

This formula can be applied for 0ºC ≤ $T$ ≤ 100°C. The speed of sound (in m/s) in water can be described by (Del Grosso & Mader 1972, Lubbers & Graaff 1998)

$$s_{Water} = 24{,}130\left(0.1983\alpha^5 - 1.3322\alpha^4 + 3.6287\alpha^3 - 5.0548\alpha^2 + 3.6182\alpha - 1\right). \tag{3.42}$$

This relation is applicable for 0ºC ≤ $T$ ≤ 100°C. The dependence of $\nu$ and $s$ on the temperature $T$ is illustrated in Fig. 3.2. The plot of the corresponding ratios in Figs. 3.2b and 3.2d shows that in particular the viscosities of air and water differ significantly. It is worth noting that the corresponding ratio of dynamic viscosities is very different to $\nu_{Air}/\nu_{Water}$ shown in Fig. 3.2 because we have $\mu_{Air} < \mu_{Water}$.

**Realization of Design Conditions.** Let us consider now the question of how it is possible to satisfy the design conditions (3.38). For doing this we consider the following case. In reality, we have to consider air at a temperature $T$ = –40ºC. Hence, we have $\nu^R$ = 10.3760 $10^{-6}$ $m^2/s$ and and $s^R$ = 306.2673 m/s. To end up with a model that is relatively small we need a model viscosity that is relatively small: see the second condition (3.38). Let us say we use water at a temperature $T$ = 40ºC. Then, we have $\nu^M$ = 0.6583 $10^{-6}$ $m^2/s$ and and $s^M$ = 1532.5990 m/s. The design conditions (3.38) read for these values

$$\frac{v^M}{v^R} = 5.0041, \qquad \frac{{d_2}^M}{{d_2}^R} = \frac{1}{78.8737}. \tag{3.43}$$

Hence, a complete similarity can be achieved by a model system that is about 79 times smaller than the real system and a model velocity that is five times higher than the real velocity. It may be the case that there is no way to realize a model velocity that is higher than the real velocity. This problem can be solved if the aircraft velocity is much smaller than the velocity of sound $s^R$ = 306.2673 m/s.

For this case, the lift coefficient can be considered to be independent of $s$. Then, the design conditions (3.37) in addition to the conditions $\theta^M = \theta^R$ and $(d_1 / d_2)^M = (d_1 / d_2)^R$ reduce to only one condition,

$$\frac{v^M}{v^R} = \frac{\nu^M}{\nu^R} \frac{{d_2}^R}{{d_2}^M}. \tag{3.44}$$

For the conditions considered, $\nu^R = 10.3760\ 10^{-6}$ m$^2$/s and $\nu^M = 0.6583\ 10^{-6}$ m$^2$/s, this relation is given by

$$\frac{v^M}{v^R} = \frac{1}{15.7618} \frac{{d_2}^R}{{d_2}^M}. \tag{3.45}$$

For the condition $v^M = v^R$ that we can realize we find ${d_2}^M \approx {d_2}^R / 16$, this means the model system has to be about 16 times smaller than the real system.

## 3.3 Applications of Low-Complexity

We will consider three examples to illustrate the use of dimensional analysis: the vehicular stopping distance and Kepler's Third Law, which were considered in Chap. 1, and Stokes' Law, which will be applied in Chap. 6. The problems considered here have a low complexity, which means that the nondimensional equations to be obtained will involve only one nondimensional product.

### *3.3.1 Vehicular Stopping Distance*

**Two Processes.** Let us consider the use of dimensional analysis regarding the modeling of the total vehicular stopping distance that was discussed in Sect. 1.2.3. We have to define first the variables that determine this problem. There are two processes that contribute to the total stopping distance $D$,

$$D = D_R + D_B. \tag{3.46}$$

The first contribution $D_R$ refers to the reaction distance. The reaction distance is the vehicular travel distance between two events: the moment at which the driver realizes the need to stop the vehicle, and the moment at which the brakes are actually applied. The second contribution to the total stopping distance $D$ is the braking distance $D_B$. The braking distance is the vehicular travel distance between the moment at which the brakes are actually applied, and the moment at which the vehicle stops. Let us separately analyze these two processes.

**Reaction Distance.** Which variables will determine the reaction distance? First, the reaction distance will depend on the driver's reaction time $T_R$: the faster the reaction time, the smaller will be the reaction distance. Second, the reaction distance will depend on the vehicular velocity $v$: the higher the velocity, the larger will be the reaction distance. Therefore, we can expect the reaction distance as a function $D_R = D_R(T_R, v)$. The dimensional constraint reads then

$$D_R{}^a T_R{}^b v^c = c_1, \tag{3.47}$$

where $c_1$ is any constant that is independent of $L$, $M$, and $T$. By accounting for the dimensions of $D_R$, $T_R$, and $v$, Eq. (3.47) implies that

$$L^a T^b \left(\frac{L}{T}\right)^c = L^{a+c} T^{b-c} = c_2, \tag{3.48}$$

where $c_2$ is a constant. The condition that the left-hand side is independent of $L$, $M$, and $T$ then leads to the following two conditions (there is no third condition because $M$ is not involved here),

$$0 = a + c, \tag{3.49a}$$

$$0 = b - c. \tag{3.49b}$$

We use the exponent $a$ as an independent variable because we would like to solve for $D_R$. Hence, $c = -a$, and $b = c = -a$. Equation (3.47) can be written then

$$D_R{}^a T_R{}^{-a} v^{-a} = \left(\frac{D_R}{T_R v}\right) = c_1. \tag{3.50}$$

We obtain, therefore, only one nondimensional product $D_R / (T_R v)$. This product must be constant because there is no other nondimensional product that can affect $D_R / (T_R v)$. Correspondingly,

$$D_R = c_R T_R v, \tag{3.51}$$

where the constant $c_R$ is introduced. We may set $c_R = 1$. In this case, the formula obtained represents the usual formula for the distance traveled by a body that moves with a constant velocity,

$$D_R = T_R v. \tag{3.52}$$

**Braking Distance.** Which variables will determine the braking distance $D_B$? The braking distance is determined by the brake force $F_B$ applied. It will also depend on the vehicular velocity $v$: the higher the velocity, the longer will be the braking distance. The mass $m$ of the vehicle considered will also influence the braking distance: the higher the vehicular mass, the longer will it take to stop. Therefore, we can expect the braking distance as the function $D_B = D_B(F_B, v, m)$.

We set up the constraint for the dimensional analysis,

$$D_B{}^a \, F_B{}^b \, v^c \, m^d = c_1, \tag{3.53}$$

where $c_1$ is any constant. By accounting for the dimensions of $D_B$, $F_B$, $v$, and $m$, the latter condition leads to

$$L^a \left( \frac{M\,L}{T^2} \right)^b \left( \frac{L}{T} \right)^c M^d = L^{a+b+c} \, M^{b+d} \, T^{-2b-c} = c_2, \tag{3.54}$$

where $c_2$ is a constant. The condition that the left-hand side is independent of $L$, $M$, and $T$ then implies three conditions:

$$0 = a + b + c, \tag{3.55a}$$

$$0 = b + d, \tag{3.55b}$$

$$0 = -2b - c. \tag{3.55c}$$

For deriving a formula for $D_B$ we use the parameter $a$ as an independent variable. The latter two relations imply $b = -c/2 = -d$, such that the first relation implies $a = -b + 2\,b = b$. Correspondingly, there are the following three conditions for the dependent variables $b$, $c$, and $d$,

$$b = a, \tag{3.56a}$$

$$d = -a. \tag{3.56b}$$

$$c = -2a. \tag{3.56c}$$

Equation (3.53) can be written then

$$D_B{}^a \, F_B{}^{a} \; {}^{-2a} \, m^{-a} = \left( \frac{F_B \, D_B}{m v^2} \right)^a = c_1. \tag{3.57}$$

The result of the dimensional analysis is, therefore, one nondimensional product $F_B \, D_B / (m \, v^2)$, which must be constant, i.e.

$$D_B = c_B \frac{m v^2}{F_B}. \tag{3.58}$$

The constant $c_B$ introduced here can be calculated by considering the energy balance. The work needed to reduce the kinetic energy $m\,v^2/2$ of the car to zero is given by $F_B \, D_B$. In particular, the work $F_B \, D_B$ has to balance the kinetic energy $m\,v^2/2$, this means $F_B \, D_B = m\,v^2/2$. A look at Eq. (3.58) reveals that this requirement implies $c_B = 1/2$. Hence, the braking distance reads

$$D_B = \frac{v^2}{2\,g_B}. \tag{3.59}$$

Here, the constant braking acceleration $g_B = F_B / m$ is introduced.

**Comparison with Experiments.** Accordingly, the total stopping distance $D = D_R + D_B$ is found to be given by

$$D = \left( T_R + \frac{v}{2\,g_B} \right) v. \tag{3.60}$$

The comparison with Eq. (1.9) derived from experimental data,

$$D = \left( 2 + \frac{v}{20} \right) v, \tag{3.61}$$

reveals the validity of the conclusion (3.60) obtained here: $D$ is the sum of a linear and a quadratic function of $v$. The comparison of Eqs. (3.60) and (3.61) enables the calculation of the unknown parameters in Eq. (3.60). In Eq. (3.61), the total stopping distance $D$ is measured in feet, and the vehicular velocity $v$ is measured in mph. Hence, $T_R = 2$ ft / mph and $g_B = 10$ mph$^2$ / ft. By relating ft to m (1 ft = 0.3048 m) and mph to m / s (1 mph = 0.4470 m / s), the parameter values obtained are given by $T_R = 1.36$ s and $g_B = 6.56$ m / s$^2$. A mean reaction time $T_R = 1.36$ s appears to be a reasonable value. A braking acceleration $g_B$ equal to 67% of the gravity acceleration is also a plausible value.

### *3.3.2 Kepler's Third Law*

**Dimensional Analysis.** Next, let us consider Kepler's Third Law discussed in Sect. 1.2.2 as another example. What will be the relevant variables for the calculation of the orbital period $T_P$? Definitely, the orbital period will depend on the mean distance $r$ from the Sun. $T_P$ will also depend on the gravity force $F_G$ and the mass $m$ of the planet considered. Thus, we expect a function $T_P = T_P(r, F_G, m)$ for the orbital period. To set up the dimensional analysis we use the condition

$$T_P{}^a\, r^b\, F_G{}^c\, m^d = c_1, \tag{3.62}$$

where $c_1$ is any constant. By accounting for the dimensions of $T_P$, $r$, $F_G$, and $m$, this condition reads

$$T^a\, L^b \left( \frac{M\,L}{T^2} \right)^c M^d = L^{b+c}\, M^{c+d}\, T^{a-2c} = c_2, \tag{3.63}$$

where $c_2$ is any constant. This condition is satisfied if

$$0 = b + c, \tag{3.64a}$$

$$0 = c + d, \tag{3.64b}$$

$$0 = a - 2c. \tag{3.64c}$$

We use the parameter $a$ as the independent variable. The latter three conditions provide then for $b$, $c$, and $d$ the constraints

$$b = -a/2, \tag{3.65a}$$

$$d = -a/2, \tag{3.65b}$$

$$c = a/2. \tag{3.65c}$$

Hence, Eq. (3.62) can be written

$$T_P{}^a\, r^{-a/2}\, F_G{}^{a/2}\, m^{-a/2} = \left(\frac{F_G\, T_P{}^2}{m\, r}\right)^{a/2} = c_1. \tag{3.66}$$

Dimensional analysis reveals, therefore, that there is only one nondimensional product $F_G\, T_P{}^2/(m\, r)$, which must be a constant. Correspondingly, we can have

$$T_P = 2\pi\, c_P \sqrt{\frac{m\, r}{F_G}}. \tag{3.67}$$

Here, a constant $2\,\pi\, c_P$ is introduced, where $c_P$ is unknown. This formula does not involve details of the elliptical path of planets (only the mean distance $r$ from the Sun is involved). Therefore, we may assume that the planets revolve around the Sun on a circular orbit. According to the laws of mechanics we have $F_G = m\, a_c$, where the circular acceleration $a_c = 4\,\pi^2\, r\,/\,T_P{}^2$. The comparison of the resulting $F_G = 4\,\pi^2\, r\, m\,/\,T_P{}^2$ with (3.67) shows that $c_P = 1$. Hence, the result of our dimensional analysis reads

$$T_P = 2\pi \sqrt{\frac{m\, r}{F_G}}. \tag{3.68}$$

Here, the appearance of $2\,\pi$ corresponds to the usual relation between any period $T$ and the related angular frequency $\omega = 2\,\pi/\,T$. Hence, Eq. (3.68) defines the related angular frequency as $\omega_P = [F_G/(m\, r)]^{1/2}$.

**Comparison with Experiments.** Let us compare the latter formula with Eq. (1.8) obtained by the analysis of experimental data,

$$T_P = 2\pi \sqrt{\frac{r^3}{G_S}}. \tag{3.69}$$

Here, $G_S = 1.3291 \times 10^{20}\ \mathrm{m}^3/\mathrm{s}^2$ refers to the standard gravitational parameter. Both formulas for $T_P$ are different. The consistency condition between both formulas is

$$F_G = \frac{G_S m}{r^2}. \tag{3.70}$$

This relation represents Newton's Law of Gravitation if $G_S$ is written $G_S = G\,M$. Here, $G$ refers to the gravitational constant, and $M$ refers to the mass of the Sun. Therefore, Kepler's Third Law can be used for the derivation of Newton's Law of Gravitation, and vice versa. Measurements of $G$, see Gillies (1997), revealed that $G = 6.6743 \times 10^{-11}\ \mathrm{m^3}/\ (\mathrm{kg\ s^2})$. By using $G_S = 1.3291 \times 10^{20}\ \mathrm{m^3}\ /\ \mathrm{s^2}$ derived from experimental data in Sect. 1.2.2 we find the correct value $M = 1.99 \times 10^{30}$ kg for the mass of the Sun.

### *3.3.3 Stokes' Law*

**Damping Force.** Damping is relevant to many processes. Examples are given by Brownian motion, which is the motion of fine particles (pollen grains) that are suspended in water, the motion of a pendulum in air, and the motion of molecules in a fluid. In all these cases, the motion considered is damped by the presence of the surrounding particles of the medium, this means the surrounding particles act to reduce the velocity of objects considered. A simple formula for the damping force $F_d$, which is also called frictional force or drag force, is given by

$$F_d = -m\frac{v}{\tau}, \tag{3.71}$$

see, for example, the Brownian motion model discussed in Chap. 6. Here, $m$ is the mass of the (Brownian) particle considered, $v$ is the particle velocity, and $\tau$ is a characteristic time scale for the damping of particle motion. The basic idea of this assumption is that the damping force $F_d$ is linear in the velocity. The negative sign appears because the damping force does always act to reduce the particle velocity. However, the assumption (3.71) does not specify the time scale $\tau$. The effect of the medium considered on this time scale is not explained.

**Dimensional Analysis.** Let us calculate $\tau$ on the basis of dimensional analysis. This time scale will depend on the dynamic viscosity $\mu$ that reflects the influence of the medium on the damping: the higher the viscosity, the smaller will be the damping time. The damping time will also depend on the particle mass $m$: the higher the mass, the longer it will take to damp out the motion. The damping time will also depend on the size of the particle considered: the bigger the particle, the smaller will be the damping time. For simplicity, we consider a spherical particle that is characterized by its particle radius $r$. Hence, we have to expect $\tau$ as a function $\tau = \tau(\mu, m, r)$. To use dimensional analysis we consider the constraint

$$\tau^a\,\mu^b\,m^c\,r^d = c_1, \tag{3.72}$$

where $c_1$ is any constant. According to the dimensions of $\tau$, $\mu$, $m$, and $r$ (see

Table 3.1), this condition implies

$$T^a \left( \frac{M}{LT} \right)^b M^c L^d = L^{d-b} M^{b+c} T^{a-b} = c_2, \tag{3.73}$$

where $c_2$ is any constant. The left-hand side is independent of $L$, $M$, and $T$ if the following three conditions are satisfied,

$$0 = d - b, \tag{3.74a}$$

$$0 = b + c, \tag{3.74b}$$

$$0 = a - b. \tag{3.74c}$$

We use the exponent $a$ as an independent variable. The latter three conditions provide then the constraints

$$d = b = a, \tag{3.75a}$$

$$c = -b = -a, \tag{3.75b}$$

$$b = a, \tag{3.75c}$$

where the last expression was used to replace $b$ in the other two relations. Hence, Eq. (3.72) can be written

$$\tau^a \mu^a m^{-a} r^a = \left( \frac{\tau \mu r}{m} \right)^a = c_1. \tag{3.76}$$

Thus, by introducing the unknown damping constant $c_d$ we find for $\tau$ the relation

$$\tau = c_d \frac{m}{\mu r}. \tag{3.77}$$

**Comparison with Other Observations.** The constant $c_d$ can be calculated by the consideration of consequences of the Navier-Stokes equations (which will be presented in Chap. 10). This analysis was performed by Stokes (1851). He found the expression $F_d = -6 \pi \mu r v$ for the damping force regarding the motion of spherical objects in fluids with a very small Reynolds number (for laminar flows). The result $F_d = -6 \pi \mu r v$ confirms the assumption (3.71) that the damping force is a linear function of velocity. In addition, Stokes' result provides the damping time scale: the combination of $\tau = -m v / F_d$ with $F_d = -6 \pi \mu r v$ implies

$$\tau = \frac{m}{6 \pi \mu r}. \tag{3.78}$$

Hence, the damping constant is found to be $c_d = 1 / (6 \pi)$. Stokes' theoretical result is supported by measurements (Millikan 1910, 1917, 1923). The extension of Stokes' result to flows that are not characterized by a very small Reynolds number is provided by the Stokes-Cunningham formula (Li & Wang 2003).

## 3.4 Applications of Medium-Complexity: Time Measurement

Let us consider now another example for the use of dimensional analysis: the measurement of time by a pendulum. This problem has a higher complexity than the problems discussed in Sect. 3.3: we will find a nondimensional equation that involves three nondimensional products. There is then the question of how is it possible to make the result of dimensional analysis useful, this means how is it possible to obtain an explicit formula for the definition of time on this basis.

### *3.4.1 The Pendulum Period*

**Time Measurements.** The measurement of time is clearly relevant to our life: the specification of time is helpful to make appointments, time is used to analyze many processes (like weather or climate), and time is used to define other relevant variables (like velocity). Thus there is a long history of the measurement of time: sundials, water, candle, sand, and mechanical clocks were applied over centuries. Inspired by investigations of pendulums by Galileo Galilei around 1602, the pendulum clock was invented and patented by the Dutch scientist Christiaan Huygens in 1656. Galileo discovered the key property that makes pendulums useful timekeepers: isochronism, which means that the period of swing of a pendulum is approximately the same for differently sized swings. Huygens' pendulum clock was more accurate than any other measurement methods used before: his clock had an error of less than one minute per day. His later refinements reduced the clock's errors to less than 10 seconds per day. From its invention by Huygens until the 1930s, the pendulum clock was the World's most accurate timekeeper. During the Industrial Revolution (1700–1900), daily life was organized around the home pendulum clock. More accurate pendulum clocks, called regulators, were installed in places of business and used to schedule work and set other clocks. The most accurate pendulum clocks, known as astronomical regulators, were used in observatories for astronomy, surveying, and for celestial navigation. Beginning in the 1800s, astronomical regulators in naval observatories served as primary standards for national time distribution services.

**Pendulum.** An illustration of a pendulum is given in Fig. 3.3. A mass $m$ is attached to one end of a rigid, but weightless, supported rod of length $r$. The rod is free to rotate in one plane. The angle $\alpha_0$ refers to the initial angle of displacement from the vertical. The pendulum is driven by the gravity force that is accounted for in Fig. 3.3 by the gravity acceleration $g$. Another force that may affect the pendulum motion is given by the damping force due to the air resistance. The damping force reduces gradually the amplitude of oscillations (see Fig. 3.3b).

a) Undamped Pendulum b) Damped Pendulum

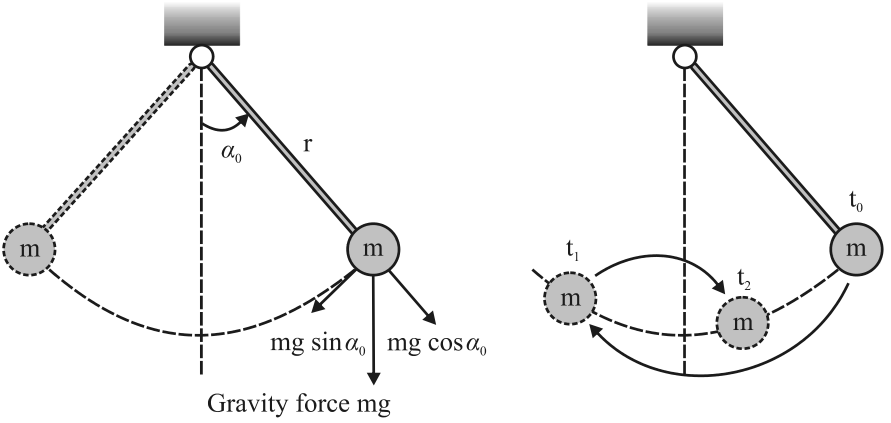



**Fig. 3.3.** A pendulum. The left-hand-side figure illustrates an undamped pendulum that swings between two maximum positions at the same angle. The right-hand side figure illustrates a damped pendulum where the damping reduces the amplitude of oscillations.

The damping force is caused by the fluid viscosity $\mu$. In contrast to the gravity force there is no exact knowledge available about the structure of the damping force. First of all, we would like to know how this force depends on the velocity (whether it is, e.g., a linear or quadratic function). Unfortunately, this dependence of the damping force on the velocity is often unknown. Usually, the effect of the damping force is involved by modeling this force according to Stokes' Law (3.71): see Sect. 9.4.

### *3.4.2 Dimensional Analysis*

**Pendulum Period.** The most important variable of the pendulum problem is its period $T_P$, which is the time required for the pendulum bob to swing through one complete cycle and return to its original position. The relevance of $T_P$ is given by the fact that $T_P$ determines the unit for the measurement of time. To calculate $T_P$, we have to know which variables may affect $T_P$. According to the discussion in Sect. 3.4.1 we expect that $T_P$ depends on the pendulum mass $m$, length $r$, and initial angle of displacement $\alpha_0$. $T_P$ may also depend on the gravity acceleration $g$ and viscosity $\mu$. Thus, we expect $T_P$ as a function of the following parameters,

$$T_P = T_P(\alpha_0, r, g, m, \mu). \tag{3.79}$$

It is worth noting that the damped pendulum motion is not periodic because of the amplitude reduction. Nevertheless, the pendulum period $T_P$ still can be used to characterize the pendulum motion. $T_P$ is defined for this case as the time required for the mass to swing once forth and back (the time between $t_0$ and $t_2$ in Fig. 3.3).

**Dimensional Analysis.** According to Eq. (3.79), the requirement for a dimensionally correct relation for the pendulum period $T_P$ is given by

$$T_P{}^a\, \alpha_0{}^b\, r^c\, g^d\, m^e\, \mu^f = c_1, \tag{3.80}$$

where $c_1$ is any constant. By taking the dimensions of $T_P$, $\alpha_0$, $r$, $g$, $m$, and $\mu$ into account, Eq. (3.80) reads

$$T^a \left(L^0\, M^0\, T^0\right)^b L^c \left(\frac{L}{T^2}\right)^d M^e \left(\frac{M}{LT}\right)^f = L^{c+d-f}\, M^{e+f}\, T^{a-2d-f} = c_2, \tag{3.81}$$

where $c_2$ is any constant. The condition that the left-hand side is independent of $L$, $M$, and $T$ then implies the following three conditions,

$$0 = c + d - f, \tag{3.82a}$$

$$0 = e + f, \tag{3.82b}$$

$$0 = a - 2d - f. \tag{3.82c}$$

We use the exponent $a$ as an independent variable because we like to calculate $T_P$. In addition to $a$, we use $f$ as an independent variable to keep the damping effect separated from the other variables (this means we would like to have a formula for the pendulum period where $\mu$ appears only ones). Equations (3.82) provide then for $c$, $d$, and $e$ the constraints

$$c = -\frac{a-f}{2} + f = \frac{3f-a}{2}, \tag{3.83a}$$

$$e = -f, \tag{3.83b}$$

$$d = \frac{a-f}{2}. \tag{3.83c}$$

Here, Eq. (3.83c) was used to obtain Eq. (3.83a). Accordingly, Eq. (3.80) can be written

$$T_P{}^a\, \alpha_0{}^b\, r^{3f/2-a/2}\, g^{a/2-f/2}\, m^{-f}\, \mu^f = \left(\frac{T_P}{\sqrt{r/g}}\right)^a \alpha_0{}^b \left(\frac{\mu}{m}\sqrt{\frac{r^3}{g}}\right)^f = c_1. \tag{3.84}$$

Consequently, there are three nondimensional products that can be related to each other: $T_P / (r/g)^{1/2}$, $\alpha_0$, and $\mu\, (r^3/g)^{1/2} / m$. Hence, the pendulum period $T_P$ can be written

$$T_P = 2\pi\, c_P \sqrt{\frac{r}{g}}. \tag{3.85}$$

Here, a factor $2\pi\, c_P$ is introduced. The factor $2\,\pi$ accounts for the usual relation between the period $T_P$ and the related angular frequency $\omega_P = 2\,\pi/T_P$. The nondimensional product $c_P$ in Eq. (3.85) can be a function of $\alpha_0$ and $\mu\,(r^3/g)^{1/2}/m$,

$$c_P = c_P(\alpha_0, \mu_*). \tag{3.86}$$

The nondimensional variable $\mu_* = \mu\,(r^3/g)^{1/2}/m$ is introduced here as an abbreviation to simplify the notation.

### *3.4.3 Comparisons with Other Results*

**Nonlinear Pendulum Equation.** To use the pendulum period relation (3.85) for calculations of $T_P$ we need to know how the coefficient $c_P$ depends on $\alpha_0$ and the damping variable $\mu_*$. One way to address this problem would be the attempt to use measurements for the investigation of $c_P = c_P(\alpha_0, \mu_*)$. This approach is rather expensive, and it is affected by randomness. A much more appropriate approach is given by the use of the pendulum equation for the calculation of $c_P$. The latter equation is a consequence of Newton's Second Law. By defining the gravity and damping forces, the equation for the pendulum written in terms of nondimensional parameters reads (see Sect. 9.4),

$$\frac{d^2\alpha}{dt_*^{\,2}} = -d\,\mu_*\frac{d\alpha}{dt_*} - \sin\alpha. \tag{3.87}$$

Here, $\alpha = \alpha(t)$ is the pendulum angle that changes in time, and $t_* = t/(r/g)^{1/2}$ is a nondimensional time. The two contributions on the right-hand side of Eq. (3.87) account for the effect of gravity (the last term) and damping (the term involving $\mu_*$). The factor $d$ refers to a nondimensional parameter that depends, for example, on the fluid considered and the properties of the pendulum bob. By adopting Stokes' Law of Friction, it turns out that $d = 6\,\pi\, r_P/r$ where $r_P$ refers to the radius of the pendulum bob. However, the validity of Stokes' Law is questionable in many cases, such that $d$ is considered here simply as any unknown parameter. The solution of equation (3.87) requires the specification of initial values. These initial values are given by $\alpha(t = 0) = \alpha_0$ and $d\alpha/dt_*(t = 0) = 0$. Unfortunately, Eq. (3.87) cannot be solved analytically because of the nonlinearity (the sine function) involved.

**Linear Pendulum Equation and its Solution.** However, for a relatively small initial angle of displacement $\alpha_0$ it is possible to solve Eq. (3.87). For this case, Eq. (3.87) reads

$$\frac{d^2\alpha}{dt_*^{\,2}} = -d\,\mu_*\frac{d\alpha}{dt_*} - \alpha, \tag{3.88}$$

because sin $\alpha$ can be approximated by $\alpha$. As shown in Sect. 9.4, the solution of Eq. (3.88) combined with the initial conditions considered is given by

$$\alpha = \alpha_0 \frac{e^{-d\,\mu_* t_*/2}}{\cos\delta} \cos\left( \frac{2\pi t_*}{T_P^{(LIN)}/\sqrt{r/g}} - \delta \right), \tag{3.89}$$

if the condition $d\,\mu_* < 2$ is satisfied. The latter condition simply means that the damping should be not too strong. In Eq. (3.89), $\delta$ is defined by

$$\delta = \arctan\left( \frac{d\,\mu_*/2}{\sqrt{1-(d\,\mu_*/2)^2}} \right), \tag{3.90}$$

and $T_P^{(LIN)}$ is given by the expression

$$T_P^{(LIN)} = \frac{2\pi}{\sqrt{1-(d\,\mu_*/2)^2}} \sqrt{\frac{r}{g}}. \tag{3.91}$$

The superscript "(*LIN*)" here refers to the value of $T_P$ provided by the solution (3.89) of the linear equation. According to its definition $c_P = T_P/[2\,\pi(r/g)^{1/2}]$, the coefficient $c_P$ is then given by

$$c_P^{(LIN)} = \frac{1}{\sqrt{1-(d\,\mu_*/2)^2}}. \tag{3.92}$$

It is worth noting that the solution (3.89) of the linear pendulum equation (3.88) implies that $c_P^{(LIN)}$ is independent of the initial angle of displacement $\alpha_0$. To prepare the discussion below we also use the solution (3.89) of the linear equation for the calculation of the ratio

$$\frac{\alpha_2^{(LIN)}}{\alpha_0} = \exp\left\{ -\frac{\pi\, d\,\mu_*}{\sqrt{1-(d\,\mu_*/2)^2}} \right\}. \tag{3.93}$$

Here, $\alpha_2^{(LIN)}$ is the angle at the time $t_2$ (see Fig. 3.3) calculated by the solution of the linear pendulum equation. The ratio $\alpha_2^{(LIN)}/\alpha_0$ describes the reduction of the maximum angle due to damping. The linear pendulum equation implies that the ratio $\alpha_2^{(LIN)}/\alpha_0$ is found to be independent of the initial angle of displacement $\alpha_0$.

**Range of Damping Considered.** To see the range of applicability of the linear Eq. (3.88) and its consequences we have to solve numerically the nonlinear Eq. (3.87). This requires an appropriate choice of the $d\,\mu_*$ range considered. The latter question is addressed in terms of Fig. 3.4a that shows the angle reduction due to damping after the first two swings for $d\,\mu_*$ variations $0.01 \le d\,\mu_* \le 1$. It may be seen that the consequence $\alpha_2^{(LIN)}/\alpha_0$ of the linear pendulum equation, which is

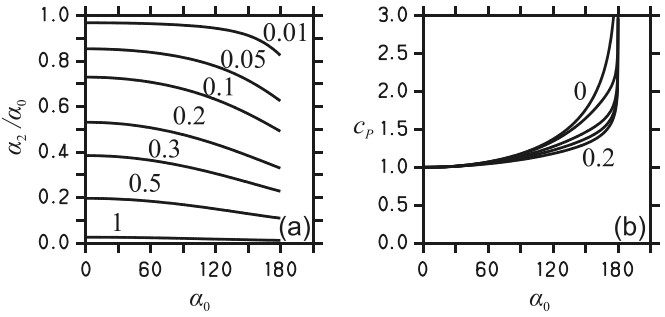



**Fig. 3.4.** The nonlinear damped pendulum. $\alpha_2 / \alpha_0$ is shown in (a) in dependence on the initial angle of displacement $\alpha_0$. Here, $\alpha_2$ is the angle at the time $t_2$ (see Fig. 3.3). The values on the curves refer to the $d\,\mu_*$ values applied. (b) shows $c_P$ for the nonlinear pendulum in dependence on the initial angle of displacement $\alpha_0$. The $d\,\mu_*$ values applied are given by 0, 0.01, 0.05, 0.1, and 0.2, respectively.

independent of $\alpha_0$, is applicable for relatively small $\alpha_0$ values. We also observe that relatively small damping values $d\,\mu_*$ already result in a significant maximum angle reduction. For example, for $d\,\mu_* = 0.2$ we find $\alpha_2 / \alpha_0 = 0.532$ for very small $\alpha_0$ values. Thus, we will only consider a range $d\,\mu_* \le 0.2$ in the following to allow the pendulum at least a few swings. For higher damping values it turns out that the concept to calculate the pendulum period $T_P$ for the damped pendulum becomes questionable: it is hardly possible to talk about a pendulum period if the pendulum motion disappears after a few swings.

**Nonlinear Damped Pendulum Period.** The pendulum period of the nonlinear damped pendulum can be calculated by solving the nonlinear pendulum equation (3.87) numerically. The results obtained for the range $0 \le d\,\mu_* \le 0.2$ of $d\,\mu_*$ values are shown in Fig. 3.4b. For relatively small $\alpha_0$ the $c_P$ values correspond to the result (3.92) of the linear pendulum equation, which is independent of $\alpha_0$. All the $c_P^{(LIN)}$ values are found to be very close to 1 because of the range of $d\,\mu_*$ considered. For example, we have $c_P^{(LIN)} = 1.005$ for the highest damping value $d\,\mu_* = 0.2$ applied. The $c_P$ value for the undamped ($d\,\mu_* = 0$) nonlinear pendulum shows a significant increase of $c_P$ values for initial angles $\alpha_0 \ge 90°$. The reason for this increase is the following. The sin $\alpha$ function is always smaller than $\alpha$. Thus, the gravity force $m\,g$ sin $\alpha$ that drives the pendulum motion is smaller for the nonlinear case than for the linear approximation. The smaller driving force of the pendulum is then related to a longer time required to swing once forth and back, i.e., a higher $c_P$ and $T_P$. The effect of damping on $c_P$ for the range $\alpha_0 \ge 90°$ is such that the $c_P$ values are smaller than for the undamped pendulum. An explanation for this finding is given by the reduction of the maximum angle values due to damping. The smaller angle variation reduces the time required for the first two swings of the pendulum.

**Table 3.2** The $c_P$ values of the damped nonlinear pendulum in dependence on the initial angle $\alpha_0$ and damping value $d\,\mu_*$.

| $d\,\mu_*$ | $\alpha_0 = 5°$ | $\alpha_0 = 10°$ | $\alpha_0 = 15°$ | $\alpha_0 = 20°$ | $\alpha_0 = 25°$ | $\alpha_0 = 30°$ |
|---|---|---|---|---|---|---|
| 0.00 | 1.0005 | 1.0019 | 1.0043 | 1.0077 | 1.0120 | 1.0174 |
| 0.01 | 1.0005 | 1.0019 | 1.0042 | 1.0074 | 1.0117 | 1.0169 |
| 0.05 | 1.0007 | 1.0020 | 1.0040 | 1.0069 | 1.0106 | 1.0152 |
| 0.10 | 1.0016 | 1.0027 | 1.0045 | 1.0070 | 1.0103 | 1.0143 |
| 0.20 | 1.0053 | 1.0062 | 1.0077 | 1.0097 | 1.0123 | 1.0155 |

**Seconds Pendulum.** A look at the $c_P$ values shown in Table 3.2 for various initial angles of displacement and damping values shows that $c_P = 1$ represents a very good approximation if the damping and initial angle of displacement are not too high. For the damping values applied one finds that the deviations of $c_P$ values from 1 are smaller than 1% if $\alpha_0 \leq 20°$. Thus, the pendulum period is given by

$$T_P = 2\pi\sqrt{\frac{r}{g}} \tag{3.94}$$

for $\alpha_0 \leq 20°$. This formula reveals the requirements for the design of a seconds pendulum, which was used for the definition of time after 1670. A seconds pendulum is a pendulum whose period is precisely two seconds, one second for a swing forward and one second for a swing backward. By setting $T_P = 2$ s and using $g = 9.81\ \mathrm{m/s^2}$ for the gravity acceleration, we find the requirement $r = 0.9940$ m, this means $r$ has to be about 1 m. It is worth noting that the seconds pendulum can be also applied for the definition of length. In 1790, Talleyrand proposed that the meter is the length of the seconds pendulum at a latitude of 45°.

## 3.5 Applications of High-Complexity: Lift

After demonstrating the value of dimensional analysis in Sects. 3.3 and 3.4, let us come back to the lift problem considered in the beginning. This problem has a higher complexity than the problems discussed before: it involves a nondimensional equation that relates five nondimensional products. The focus here is not on the use of dimensional analysis (the nondimensional equation for this problem was derived in Sect. 3.2), but on the following. The nondimensional equation provided by dimensional analysis is not really helpful without having a method to close this equation by the determination of the lift coefficient: the lift coefficient has to be known to calculate the lift force. The determination of coefficients can be a hard task if the coefficient is a function of many parameters (as given for the lift coefficient that is a function of four parameters). This determination usually

requires solutions for four problems: (1) the understanding of the meaning of parameters involved, (2) the understanding of the range of parameter variations that have be considered, (3) the finding of a basic solution (the determination of the influence of the process-controlling parameter), and (4) insight into modifications of the influence of the process-controlling parameter due to variations of other parameters. This process will be illustrated here regarding the relevant lift force problem (a 1% increase in maximum lift coefficient is equivalent to a 2000 kg increase in payload at a fixed approach speed).

### *3.5.1 The Problem*

**Lift Coefficient Formula.** To prepare the following discussion let us have a closer look at the meaning of parameters involved in the lift force calculation. According to Eq. (3.20), the lift force $F_L = C_L \rho v^2 d_2^2 / 2$ depends on the lift coefficient $C_L$ that is given by

$$C_L = C_L\left(\theta, \frac{d_1}{d_2}, Ma, Re\right). \tag{3.95}$$

Here, the Mach number $Ma$ is defined by

$$Ma = \frac{v}{s}, \tag{3.96}$$

and the Reynolds number $Re$ is defined by

$$Re = \frac{\rho d_2 v}{\mu} = \frac{d_2 v}{\nu}. \tag{3.97}$$

The last expression results from the definition of the kinematic viscosity $\nu = \mu / \rho$. Equation (3.97) represents one way to introduce the Reynolds number. Another possibility would be given by the use of $d_1$ instead of $d_2$ in Eq. (3.97). In Eq. (3.95), the lift coefficient is written as a function of $Re$ instead of a function of $Re^{-1}$. This writing does not make any difference because the dependence of $C_L$ on $Re$ is unknown anyway.

**Mach Number.** What is the meaning of the Mach number? The Mach number $Ma$ measures the compressibility of a fluid. When you pump up a tire you can fill it with air, and then pump in more air, with no appreciable change in volume. You cannot do the same with water. Thus, air is compressible (it has a relatively high compressibility) and water is incompressible (it has a relatively low compressibility). The high compressibility of air can affect the motion of objects in air. When you are driving along in your car at 120 km / h, air seems to be incompressible,

because when you push at such a speed, the air has plenty of time to get out of the way. At these relatively low speeds, the air behaves just like water. However, as your aircraft speeds up, and $Ma \approx 0.3$, the air molecules cannot get out of the way fast enough, so there are the first signs of changes in density in the air around the aircraft. Obviously, for higher Mach numbers there is a much stronger increase in density in the air around the aircraft.

**Reynolds Number.** What is the meaning of the Reynolds number? The inverse Reynolds number $Re^{-1}$ represents a measure for the viscosity of a fluid. Viscosity is a measure of the resistance of a fluid to be deformed or to flow. Fluids may have a very different viscosity: the dynamic viscosity of air, for example, is much smaller than the dynamic viscosity of water. The reason for that is given by the different intermolecular forces of air and water: the intermolecular forces (which hinder the flow) of water are much stronger than the intermolecular forces of air. The viscosity may have a relevant influence on the motion of objects in fluids. For the same propulsion power, the speed of the object will be higher in a fluid with a lower viscosity.

### *3.5.2 The Range of Parameter Variations*

**Angle of Attack Variation.** A reasonable range of the $\theta$ variation can be easily determined: the angle of attack should be in the range $0 \le \theta \le 45^{\circ}$.

**Shape Factor Variation.** Figure 3.5 illustrates the possible range of variations of the shape factor $d_1 / d_2$. For the NACA 0010 and NACA 0020 airfoils, the ratios of maximum width to maximum length are given by 0.1 and 0.2, respectively. These ratios can be used to characterize the ratio $d_1 / d_2$. Thus, a reasonable range of $d_1 / d_2$ variations is given by $0.1 \le d_1 / d_2 \le 0.2$.

**Mach Number Variation.** To determine the range of variations of the Mach number $Ma = v / s$ we need to specify the variation of the velocity of sound $s$ and the relative aircraft velocity $v$. For dry air at standard atmospheric pressure, $s$ is given by Eq. (3.40). To cover a range of sound velocity variations we consider $s$ at 20ºC (where $s = 343.4$ m / s = 1236 km / h) and $s$ at –40º (where $s = 306.3$ m / s = 1103 km / h). The corresponding curves for the Mach number are shown in Fig. 3.6 for velocities up to 2000 km / h. It may be seen that the range of Mach number variations is $0 \le Ma \le 1.8$.

**Fig. 3.5.** NACA 0010 and 0020 airfoils (*solid* and *dashed lines*, respectively).

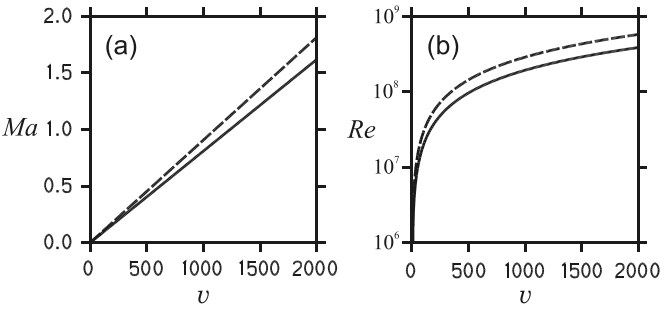



**Fig. 3.6.** The dependence of the Mach number *Ma* and Reynolds number *Re* on the velocity (in km / h). The *solid* and *dashed lines* refer to sound velocity and viscosity values that are calculated for a temperature of 20ºC and –40ºC, respectively. Here, $d_2 = 3$ m is assumed.

**Reynolds Number Variation.** The Reynolds number $Re = d_2\, v\, /\, \nu$ depends on the viscosity $\nu$ and the relative aircraft velocity $v$. For air at standard atmospheric pressure, the viscosity is given by Eq. (3.39). To have reference values, we will consider the kinematic viscosity at 20ºC (where $\nu = 1.55 \times 10^{-5}$ m$^2$ / s) and at –40ºC (where $\nu = 1.04 \times 10^{-5}$ m$^2$ / s). The corresponding curves for the Reynolds number are shown in Fig. 3.6 for velocities up to 2000 km / h, where $d_2 = 3$ m is used. It may be seen that the Reynolds number may vary over many orders of magnitude. In particular, we have to expect values $10^6 \le Re \le 10^9$. The temperature has an influence on the Reynolds number, but the effect is rather limited in comparison to the huge variations of the Reynolds number with the velocity $v$.

### *3.5.3 The Lift Coefficient Calculation*

**Lift Coefficient Calculation.** The lift coefficient can be calculated by means of the computational fluid dynamics (CFD) code XFOIL (http://web.mit.edu/drela/Public/web/xfoil/). It is worth emphasizing that this calculation does not exactly provide the lift coefficient $C_L$ as defined by $F_L = C_L\, \rho\, v^2\, d_2^2 \,/\, 2$. Usually, $C_L$ is calculated by using the wing area instead of $d_2^2$ in $F_L = C_L\, \rho\, v^2\, d_2^2 \,/\, 2$. The wing area is determined by the chord and the span length. The span length does not enter $F_L = C_L\, \rho\, v^2\, d_2^2 \,/\, 2$. However, one may expect that the span length is comparable to $d_2$. Depending on the shape of the wing, the lift coefficient calculated by the CFD code is then equal to $\beta\, C_L$, where the constant $\beta$ is of order unity. To focus on the main features, we will assume here for simplicity that $\beta = 1$.

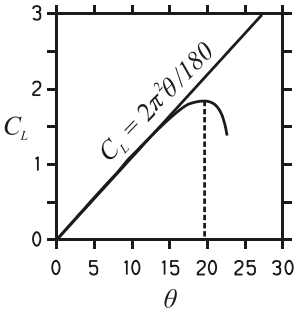



**Fig. 3.7.** The lift coefficient $C_L$ (*solid curve*) of a NACA 0010 airfoil for $Re = 10^7$ and $Ma = 0$. The stall angle indicated by the *dashed line* is 19.6º. The linear curve $C_L = 2\,\pi^2\,\theta\,/\,180$ shows the prediction of the thin airfoil theory.

**Angle of Attack Effects.** Let us understand first the basic behavior of the lift coefficient that is determined by its variation with the angle of attack $\theta$ (which is the most relevant variable). A typical lift coefficient curve obtained with the CFD code is given in Fig. 3.7. For relatively small angles $\theta$ we find a very good agreement of $C_L$ with the linear function $C_L = 2\,\pi^2\,\theta\,/\,180$ that is provided by the thin airfoil theory (Abbot & Doenhoff 1959). This linear increase is the expected behavior: more air passing under the wing is deflected downward by the bottom surface of the wing. Thus, the pressure below the airfoil increases whereas the pressure above the airfoil decreases. The vertical pressure difference results in the lift force that pushes the airplane up. But there is a limit to this, i.e., there is a maximum lift coefficient at the stall angle. For angles higher than the stall angle there are no well-organized high and low pressure areas anymore: the pressure distribution now becomes random due to the appearance of turbulence (the high-pressure air under the wing can more easily move around the back of the wing toward the low-pressure air on top of the wing, thus weakening that low-pressure area). Thus, the lift coefficient will decrease for angles higher than the stall angle.

**Mach Number Effects.** The effect of Mach number variations on the lift coefficient is shown in Fig. 3.8a regarding the NACA 0010 airfoil. An increasing Mach number reduces the lift coefficient significantly. This observation can be explained by the increasing air density around the aircraft: the air molecules cannot get out of the way fast enough. The increasing air density then decreases the vertical pressure difference that pushes the airplane up. We see a strong reduction of the stall angle, which has implications on the flight behavior.

**Reynolds Number Effects.** The effect of Reynolds number variations on the lift coefficient is shown in Fig. 3.8b for the NACA 0010 airfoil. A decrease of the Reynolds number $Re = d_2\,v\,/\,\nu$ can be seen as a reduction of the relative aircraft velocity $v$. The latter implies a reduction of the amount of air that is deflected downward, which decreases the lift. The decrease of the lift coefficient for this case looks similar to the effect of an increasing Mach number. A difference is given by the fact that the lift coefficient still follows the prediction of the thin airfoil theory for relatively small angles of attack.

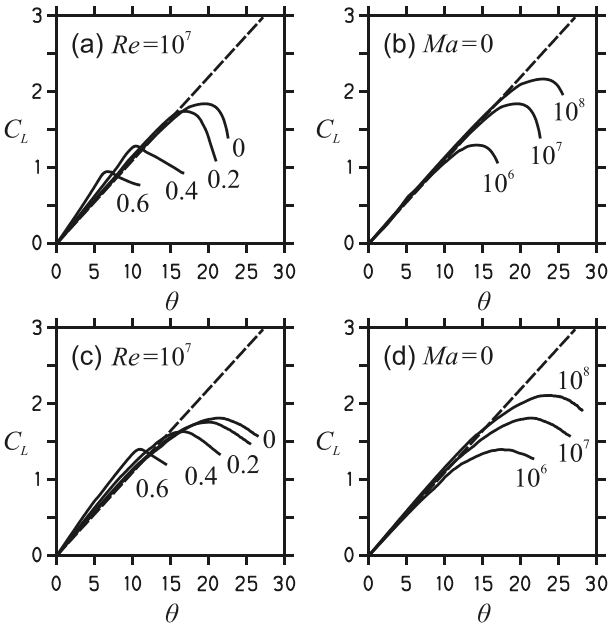



**Fig. 3.8.** Lift coefficients for the NACA 0010 airfoil (*upper pictures*: $d_{12} = 0.1$) and NACA 0020 airfoil (*lower pictures*: $d_{12} = 0.2$). In (a) and (c), the Reynolds number $Re = 10^7$ is fixed and Mach number variations are considered (the values related to the curves show $Ma$). In (b) and (d), the Mach number $Ma = 0$ is fixed and Reynolds number variations are considered (the values related to the curves show $Re$). The *dashed lines* show the thin airfoil curve $C_L = 2\,\pi^2\,\theta/\,180$.

**Shape Factor Effects.** The effect of $d_1 / d_2$ variations is given by Figs. 3.8c–d that show the effect of Mach and Reynolds number variations for the NACA 0020 airfoil. The trend of these curves is similar to the trends given in the upper pictures for the NACA 0010 airfoil. However, $d_1 / d_2$ does affect these trends: we see that a higher $d_1 / d_2$ slightly increases the deviations from the thin airfoil curve $C_L = 2\,\pi^2\,\theta/180$.

## 3.6 Summary

This chapter was organized into two parts. First, the use of dimensional analysis was explained in Sect. 3.2. Second, applications of such analytical results to problems of increasing complexity were discussed in Sects. 3.3–3.5 by combining the conclusions of dimensional analysis with the results of simulations, theory and experiments. Let us summarize the basic observations made here.

**Dimensional Analysis.** Dimensional analysis allows us to look at problems in their simplest form. Regarding the lift force, for example, we see that the lift force $F_L = F_L(v, d_1, d_2, \theta, \rho, \mu, s)$, which is a function of seven variables, cannot be any function, but only certain function types are dimensionally correct equations. Such a correct equation is $F_L = C_L \rho v^2 d_2^2 / 2$, where $C_L$ is a function of four parameters, $C_L = C_L(\theta, d_1 / d_2, Ma, Re)$. The function obtained by dimensional analysis is not unique: another correct equation would be $F_L = C_L \rho v^2 d_1^2 / 2$. The latter equation can be also written $F_L = (C_L d_1^2 / d_2^2) \rho v^2 d_2^2 / 2$, this means the difference between both equations is a modified lift coefficient. The conclusions obtained by dimensional analysis are very helpful:

- First, we can use the result of dimensional analysis to make relatively reasonable predictions without having any additional information. Coefficients like $C_L$ here are often found to be of order unity. For example, a look at Fig. 3.8 shows that usual $C_L$ variations are between zero and two. Thus, by using a value $C_L = 1$ we are already in the position to use the lift force formula for obtaining reasonable estimates.
- Second, the use of dimensional analysis is helpful for the understanding of the meaning and the relevance of parameters like the Mach and Reynolds numbers. We can understand the effects represented by these numbers (see the discussion in Sect. 3.5.1), and we can assess the conditions under which such effects can be disregarded. For example, the effect of gravity on the lift coefficient can be neglected because there are usually conditions such that $(d_2 g)^{1/2} << v$ (see the discussion in Sect. 3.1). With the same sort of reasoning we can conclude that the lift coefficient is independent of the Mach number if the relative aircraft velocity $v << s$. Similarly, we may expect that a high Reynolds number $Re \geq 10^8$ (which is often given in reality) will have little effect on the lift coefficient for a reasonable range $0 \leq \theta \leq 20^o$ of angle of attack variations. The latter view is supported by Figs. 3.8b–d. Such scale analysis represents a powerful tool to simplify equations (in addition to the problem reduction that is immediately a consequence of the use of dimensionally correct equations).
- Third, another relevant advantage of dimensional analysis is the possibility to design accurate experiments even if the real system (aircraft, ship, submarine) cannot be investigated. The key for doing this is the design of a similar model system. This concept does not only mean that the real and the model system have to be geometrically similar. The use of this concept requires the equality of all nondimensional parameters involved. Under conditions where a complete similarity cannot be established it is possible to design relatively similar model systems that enable the study of the most relevant system features. An example for the design of a relatively similar system was given at the end of Sect. 3.2.3, where the Mach number constraint was disregarded.

**Simulations and Related Theory.** Dimensional analysis does never provide complete solutions for a problem: there are always unknown constants that have to be specified. The determination of such constants requires the combination of conclusions of dimensional analysis with the results of simulations (the numerical solution of nonlinear differential equations that cannot be solved analytically), theory (analytical consequences of differential equations) and experiments. The most efficient approach to completely solve problems that are characterized by nondimensional equations with more than one nondimensional product is the combined use of simulations and theory. Why do we need both simulations and theory? The advantage of simulations is that such results can be obtained for more complex cases than the simple cases for which analytical results are available. The disadvantage of simulations is that such results are often only available in terms of plots, and the correctness of such results has to be justified. Such justification for simulation results can be provided by the agreement with analytical results, and analytical results are much easier to use and to understand. Thus, the combination of both methods is the most convenient way for problem solutions. The success of such a combined use of simulations and theory was shown by means of the discussion of the pendulum and lift problems in Sects. 3.4 and 3.5, respectively. The simulation results were given by solutions of the nonlinear damped pendulum equation and the lift coefficient calculation in terms of the CFD code (which gives an approximate solution to the Navier-Stokes equations). The simulation results were justified by their agreement with analytical conclusions of the differential equations considered (the solution of the linear damped pendulum equation and the thin airfoil curve $C_L = 2\,\pi^2\,\theta/\,180$, respectively).

**Experiments.** Experiments usually cover only a limited range of the conditions that can be covered by simulations, but they have an essential advantage: many experiments can provide information that is independent of assumptions (every assumption can be correct or not). This characteristic property of experiments is very valuable: the comparison of the results of simulations and theory with experiments represents the way to confirm the assumptions that are represented in terms of equations. For example, the pendulum period formula $T_P = 2\,\pi\,(r/g)^{1/2}$ is implied by assumptions regarding the interaction of the pendulum acceleration, gravity and damping forces. In principle, such ideas about the structure and interaction of forces can be wrong. Thus, experiments (regarding the variation of $T_P$ with $r$, the influence of the initial angle of displacement and the influence of the mass of the pendulum bob) are needed to confirm the correctness of theoretical results. For simplicity, the use of experimental results was illustrated here only for relatively simple problems that do only involve one nondimensional product. The discussion of the vehicular stopping distance, Kepler's Third Law, and Stokes' Law in Sect. 3.3 revealed the advantage of experiments.

## 3.7 Exercises

**3.2.1** Assume that the lift force $F_L$ is determined by the relative aircraft velocity $v$ and the total surface area $A$ of the tops of the wings.
a) Which additional parameter will affect $F_L$ due to dimensional reasons?
b) Use the dimension of $F_L$ and the parameters involved to find directly a relation for $F_L$ that is dimensionally correct.

**3.2.2** The position $s$ of a moving body at time $t$ is described by the function $s = a + bt + ct^2$.
a) Determine the dimension of the parameters $a$, $b$, and $c$.
b) Explain how $a$, $b$, and $c$ are related to physical properties of the problem considered.

**3.2.3** One model for the damping force $F_d$ assumes that $F_d$ is a linear function of the velocity $v$ of a moving body, $F_d = -a\,v$. Another model for $F_d$ assumes that the damping force is a quadratic function of $v$, $F_d = -b\,v^2$.
a) Determine the dimension of the parameters $a$ and $b$.
b) Explain how $a$ and $b$ can be related to properties of the moving body.

**3.2.4** Assume that the damping time $\tau$ of the motion of a particle in a fluid is a function of the particle radius $r$ and the kinematic viscosity $\nu$ of the fluid.
a) Use the dimensions of $\tau$, $r$, and $\nu$ to find directly a relation for $\tau$ that is dimensionally correct.
b) Does this relation represent a reasonable relation for $\tau$?

**3.2.5** Consider a raindrop falling from a cloud. Assume that the velocity $v$ of the raindrop depends on the gravity acceleration $g$.
a) Find one combination of $v$ and $g$ that has the dimension $T$. Find another combination of $v$ and $g$ that has the dimension $L$.
b) To find a nondimensional relation between $v$, $g$, and another variable we have to involve a variable that has the dimension $T$ or $L$. Explain which additional variable (with dimension $L$ or $T$) should be involved.
c) Combine $v$, $g$, and the additional variable in a nondimensional product.

**3.2.6** Consider the following equation for a spring-mass system (see Sect. 7.3),

$$m\frac{d^2y}{dt^2} + \gamma\frac{dy}{dt} + k\,y = 0.$$

Here, $y(t)$ is the displacement (dimension $L$) of the mass from its equilibrium position at time $t$. The constants involved are the mass $m$, the damping constant $\gamma$, and the spring constant $k$.

a) Determine the dimension of $\gamma$ and $k$ .
b) Use $m$, $k$, and $t$ to define a nondimensional time $t_*$.
c) Rewrite the equation such that only the nondimensional time $t_*$ and nondimensional parameters appear in addition to $y$.

**3.2.7** Consider the equation $P_1 = f(P_2, P_3, P_4)$ for any nondimensional products $P_1$, $P_2$, $P_3$, and $P_4$.
a) What is the problem regarding the use of this equation?
b) How is it possible to overcome this problem?
c) Given the case that the real system cannot be studied directly: what are the conditions for the design of a completely similar model system?
d) Given the case that a complete similarity cannot be achieved: how is it possible to design a relatively similar model system?

**3.2.8** Consider the lift force problem discussed in Sect. 3.2. The list of nondimensional products obtained is given by

$$\frac{F_L}{\rho v^2 d_2^{\,2}}, \qquad \theta, \qquad \frac{d_1}{d_2}, \qquad \frac{\mu}{\rho d_2 v}, \qquad \frac{s}{v}.$$

a) Use this list to identify the independent and dependent variables applied.
b) Assume that the gravity acceleration $g$ has to be involved in the set of variables considered. Which dependent variables used in the given list can be applied for finding a nondimensional product that involves $g$? Find the nondimensional product that involves $g$.
c) The result of nondimensional analysis can be written $F_L = C_L \rho v^2 d_2^2 / 2$. Generalize the function $C_L$ that follows from the list above by involving the nondimensional product that involves $g$.

**3.2.9** Consider the function $F_L = C_L \rho v^2 d_2^2 / 2$, where $C_L$ is given by the result of exercise 3.2.8.
a) What are the conditions for a completely similar model system? Account for the fact that $g$ cannot be changed.
b) Assume that $d_2^{\,M} = \alpha\, d_2^{\,R}$, where $0 \le \alpha \le 1$. Specify for each variable the ratio between the variable in the model system and the variable in the real system (e.g., $d_1^{\,M} / d_1^{\,R} = \ldots$, $v^M / v^R = \ldots$) as a function of $\alpha$.
c) Specify the latter design conditions for the case that the same fluid is used in the real system and in the model system.

**3.2.10** The period of a damped pendulum is given by $T_P = 2\pi c_P (r/g)^{1/2}$, where the factor is given by the function $c_P = c_P(\alpha_0, \mu_*)$. The damping parameter is defined by $\mu_* = \mu\, (r^3/g)^{1/2} / m$. Here, $r$ is the length of the pendulum, $g$ is the gravity acceleration, $\alpha_0$ denotes the initial angle of displacement, $\mu$ is the dynamic viscosity, and $m$ is the pendulum mass.

a) Describe an experiment for the calculation of $c_P$.
b) Derive the design conditions for a model pendulum that is 50% smaller than a real pendulum.
c) Specify the design conditions for the case that the same fluid is used.

**3.3.1** What is the power $P$ (dimension $M L^2 T^{-3}$) that is required to keep a vehicle of length $d$ and mass $m$ moving at a constant speed $v$?
a) Use dimensional analysis to calculate the power $P$.
b) Rewrite your result in terms of the kinetic energy $K = m v^2 / 2$, and the characteristic time scale $\tau = d / v$.

**3.3.2** The hydrostatic pressure of blood in humans is a part of the total blood pressure. The hydrostatic pressure $P$ (dimension $M L^{-1} T^{-2}$) is considered to depend on the blood density $\rho$, the height $h$ of the blood column between the heart and some lower point in the body, and the gravity acceleration $g$ (Giordano et al. 2003).
a) Use dimensional analysis to calculate the hydrostatic pressure $P$.
b) Pressure is defined as force per area. Rewrite your result for $P$ by taking reference to the gravity force $F_g = m g$ and an area $A$ that you have to define conveniently.

**3.3.3** For laminar flow in a pipe, the volume flow rate $q$ (dimension $L^3 T^{-1}$) is a function of the pipe radius $r$, the viscosity $\mu$ of the fluid, and the pressure drop per unit length $dp / dz$ (dimension $M L^{-2} T^{-2}$).
a) Use dimensional analysis to calculate the flow rate $q$.
b) How does $q$ change if the radius is increased by a factor of two?

**3.3.4** A star represents a liquid body that is held together by its own gravity. Stars may vibrate in several ways. The frequency $\omega$ of vibrations can be expected to depend on the radius $r$ of the star, the star's density $\rho$, and the gravitational constant $G$ (dimension $M^{-1} L^3 T^{-2}$).
a) Use dimensional analysis to calculate the frequency $\omega$ of vibrations of stars.
b) Explain the consequence of this result.

**3.3.5** The molecules of a gas interact in terms of attractive and repulsive forces. The attractive forces are considered to be negligible. The repulsive forces are considered to follow the relation $F = K r^{-n}$. In this relation, $r$ is the distance between the centers of two molecules, $n$ is any exponent, and $K$ is a constant of proportionality. The dynamic viscosity $\mu$ of the gas depends on the thermal velocity $v$ of a molecule, the mass $m$ of a molecule, and the coefficient of repulsion $K$.

a) Find the dimension of $K$.
b) Use dimensional analysis to calculate the dynamic viscosity $\mu$ of the gas.
c) The thermal velocity is related to the absolute temperature $T$ (in °K) by the $v = (k\,T/m)^{1/2}$, where $k$ is Boltzmann's constant. Use $v = (k\,T/m)^{1/2}$ to replace the $v$-dependence of the dynamic viscosity by a $T$-dependence. Determine the variation of the temperature dependence of the dynamic viscosity for variations $5 \le n \le \infty$.

**3.3.6** Consider Stokes' problem: a small sphere falling under gravity in a viscous fluid. The sphere is so small that the motion is very slow. The velocity $v$ of sphere depends on the radius $r$ of the sphere, the gravity force $F_g$, and the dynamic viscosity $\mu$ of the fluid.
a) Use dimensional analysis to calculate the sphere's velocity $v$.
b) Consider the case that the gravity force $F_g$ and damping force $F_d$ balance each other, this means we have $F_g + F_d = 0$. Assume that the damping force is given by Stokes' Law, $F_d = -6\,\pi\,\mu\,r\,v$. Use this fact to determine the unknown constant in the expression for $v$ obtained by dimensional analysis.
c) The density $\rho_s = m/V_s$ of the sphere is given in terms of the mass $m$ and volume $V_s = 4\,\pi\,r^3/3$ of the sphere. The gravity force is given by the relation $F_g = m\,g = \rho_s\,V_s\,g$, where $g$ refers to the gravity acceleration. Find Stokes' result for $v$ by using $F_g = \rho_s\,V_s\,g$ in the expression for $v$ obtained above.

**3.4.1** Consider again Stokes' problem: a small sphere falling under gravity in a viscous fluid. The sphere's velocity $v$ is assumed to depend on the sphere's radius $r$, the gravity acceleration $g$, the dynamic viscosity $\mu$ of the fluid, and the sphere's density $\rho_s$.
a) Use dimensional analysis to identify two nondimensional numbers: the Reynolds number $Re$ that is proportional to $\mu^{-1}$, and the Froude number $Fr$ that is proportional to $g^{-1}$.
b) Stokes' analysis showed that this problem can be described by only one nondimensional number: the ratio $Fr/Re$. Find the relation for $v$ that is implied by this conclusion.

**3.4.2** Consider again Stokes' problem described in exercise 3.4.1. Consider the case that $v$ depends in addition to $r$, $g$, $\mu$, and $\rho_s$ on the fluid density $\rho$.
a) Stokes' analysis showed that this problem can be described by two nondimensional numbers: the ratio $Fr/Re$ and the ratio $\rho/\rho_s$. Find the relation for $v$ that is implied by this conclusion.
b) Extend the result obtained for exercise 3.4.1 by assuming that $v$ does depend on $\rho$ and $\rho_s$ only via $\rho_s - \rho$.

**3.4.3** Consider a projectile that is fired with initial velocity $v_0$ at an angle $\alpha$ with the horizon.

a) Use dimensional analysis to calculate the projectile range $r$. You may expect that the gravity acceleration $g$ affects $r$.

b) You may expect that $r$ depends on $\alpha$ via the function $\sin(n\,\alpha)$, where $n$ is a constant. Determine the constant $n$ by considering special cases like $\alpha = 0,\ \pi/4,\ \pi/2$.

**3.4.4** Consider this case: A windmill is rotated by air flow to produce power to pump water. You have to find the power output $P$ (dimension $M\,L^2\,T^{-3}$) of the windmill. It can be expected that $P$ depends on the diameter $d$ of the windmill, the wind speed $v$, the air density $\rho$, the air viscosity $\mu$, and the rotational speed $\omega$ (dimension $T^{-1}$) of the windmill (Giordano et al. 2003).

a) Use dimensional analysis to calculate $P$. Use the exponents of $P$, $\mu$, and $\omega$ as independent parameters.

b) Provide the design condition for a model that is $q$ times smaller than the real windmill. Use the same fluid in the model system.

**3.4.5** Consider a steady laminar fluid flow through a smooth horizontal pipe. The pressure drop $\Delta p$ (dimension $M\,L^{-1}\,T^{-2}$) between two points along the pipe depends on the distance $d$ between the two points, the diameter $D$ of the pipe, the fluid density $\rho$, the fluid viscosity $\mu$, and the fluid velocity $v$.

a) Use dimensional analysis to calculate $\Delta p$. Use the exponents of $\Delta p$, $\mu$, and $d$ as independent parameters.

b) Provide the design condition for a model that is $q$ times smaller than the real system. Use the same fluid in the model system.

**3.5.1** The lift force $F_L = C_L\,\rho\,v^2\,d_2^{\,2}/2$ is given in terms of the chord length $d_2$, the relative aircraft velocity $v$, and the air density ρ. The lift coefficient can be described by the thin airfoil formula $C_L = 2\,\pi^2\,\theta/180$, where $\theta$ is the angle of attack.

a) Provide the design condition for a model that is $q$ times smaller than the real system.

b) Calculate for this case the ratio $F_L^M / F_L^R$ of the lift force $F_L^R$ in reality to the model lift force $F_L^M$. The fluid and the relative aircraft velocity $v$ are the same in reality and the model system.

**3.5.2** If a drop of liquid falls into a pool, a small column of liquid splashes out of the pool. The height $h$ of the column depends on the mass $m$ of the drop, the velocity $v$ of the drop, the surface tension $\sigma$ of the liquid (dimension $M\,T^{-2}$), the mass density $\rho$ of the liquid, the viscosity $\mu$ of the liquid, and the gravity acceleration $g$ (Langhaar 1951).

a) Use dimensional analysis to identify four nondimensional numbers: the length ratio $H$ that is proportional to $h$, the surface tension $T$ that is proportional to $\sigma$, the Reynolds number $Re$ that is proportional to $\mu^{-1}$, and the Froude number $Fr$ that is proportional to $g^{-1}$. Use the exponents of $h$, $\mu$, $g$, and $\sigma$ as independent parameters to find these numbers. Write these numbers by using the length scale $h_\rho = (m/\rho)^{1/3}$.
b) Present a formula for the calculation of $h$ by using $T$, $Fr$, $Re$, and $h_\rho$.

**3.5.3** An airplane is warming up its engine on the ground. The intensity of sound energy $U$ (dimension $M L^{-1} T^{-2}$) from the propeller at a distance $d$ ahead of the airplane depends in addition to $d$ on the diameter $D$ of the propeller, the rotational speed $\omega$ (dimension $T^{-1}$) of the propeller, the mass density $\rho$ of air, the air pressure $p$, and the air viscosity $\mu$ (Langhaar 1951).
a) Use dimensional analysis to identify four nondimensional numbers: the nondimensional intensity of sound energy $U_*$ that is proportional to $U$, the nondimensional distance $d_*$ that is proportional to $d$, the nondimensional rotational speed $\omega_*$ that is proportional to $\omega$, and the nondimensional viscosity $\mu_*$ that is proportional to $\mu$. Use the exponents of $U$, $d$, $\omega$, and $\mu$ as independent parameters to find these numbers. Write these numbers by using the velocity scale $v_\rho = (p/\rho)^{1/2}$, which is proportional to the velocity of sound.
b) Present a formula for the calculation of $U$ by using $d_*$, $\omega_*$, $\mu_*$, and p.

# 4 Stochastic States

The discussions in Chaps. 1 and 2 showed that randomness of data is relevant to actually all problems that involve observed data. For example, we have to deal with randomness to analyze the basic trend of noisy data of the global temperature increase. The way to address randomness effects in Chap. 2 was to minimize the deviations between a model and observations. First, there are questions about the validity of the approach applied: the use of the least-squares error was shown to be the most convenient choice among the three error concepts considered, but this approach of looking at the problem does not provide any sort of theoretical justification for the use of the least-squares error. Second, the problem considered (to find a model for the correlation of two variables) does only represent one particular problem among many questions related to the development of models that involve randomness. In particular, we need mathematical concepts for the analysis of stochastic states to determine the mean value of variables like the atmospheric temperature, the typical amount of randomness, and the probability for finding special events (very high or low temperatures). On this basis we are interested in modeling concepts that we can use for the description of properties of random variables. We need such analysis and modeling concepts for several interacting random variables (like the three interacting components of the atmospheric wind vector). Apart from the analysis and modeling of stochastic states at a particular time, our predominant goal is the development of methods for the analysis and modeling of stochastic processes that evolve in time, which is a requirement for the analysis and prediction of many processes in nature and technology. These questions will be successively discussed in Chaps. 4, 6, 8, and 10. By taking reference to analysis and modeling concepts for the description of several random variables, it will be shown that the specific correlation problem addressed in Chap. 2 (the finding of an optimal model for observations) can be formulated and treated in a much more convenient way (see Chap. 10).

S. Heinz, *Mathematical Modeling*, DOI 10.1007/978-3-642-20311-4_4,

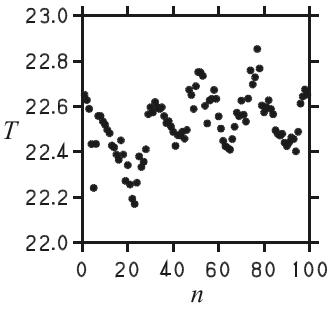



**Fig. 4.1.** An illustration of 100 measurements of the temperature $T$ (in °C) in the stably stratified atmospheric boundary layer. Here, $n$ refers to the number of the measurement value.

In preparation of the discussions in Chaps. 6, 8, and 10, this chapter addresses the analysis and modeling of properties of the state of one random variable. The requirement for the development of such analysis and modeling concepts will be explained in Sect. 4.1. Section 4.2 introduces the basic definitions used for the description of random variables. Basic methods for the modeling and analysis of properties of random variables will be presented in Sects. 4.3 and 4.4, respectively. Section 4.5 illustrates the application of these methods for the analysis and modeling of real data by considering measurements of atmospheric velocities and temperatures. The basic features of the analysis and modeling of randomness presented in this chapter will be summarized in Sect. 4.6.

It may be helpful to clarify the use of the terms random and stochastic (which are usually used with the same meaning) applied here and in the following. The term random will be used for talking about random data (the randomness of data) and random variables. The term stochastic will be used for talking about stochastic modeling, a stochastic process and a stochastic state (which means a stochastic process at a certain time).

## 4.1 Motivation

**Problem with Randomness.** We need temperature (and other) measurements to validate, e.g., the performance of weather forecasts or climate models. A typical example for such measurements is illustrated in Fig. 4.1. This figure shows 100 data points of the temperature $T$ (in °C) measured in the stationary stably stratified atmospheric boundary layer under almost the same conditions. Details about the way in which these measurements were performed are given in Sect. 4.5. Figure 4.1 indicates that it is impossible to specify only one value for the temperature $T$ under these or any other conditions (other measurements reveal the same picture). So how is it possible to make use of such random data?

**Analysis of Randomness.** The application of such data requires mathematical concepts for their analysis. First of all, we need information about the mean (also called expectation) value of random variables and the characteristic amount of randomness involved. Another often required sort of information is given by the probability for finding certain events (for example, very high or low temperature values). The specific problem regarding the calculation of such characteristics of randomness is that all such mean values are a function of the usually limited number of observations. This leads to the relevant questions of how important this dependence really is, and whether it is possible at all to arrive (for a sufficiently high number of observations) at conclusions about the statistical properties of random variables that can be reproduced by repeated experiments.

**Modeling of Randomness.** The analysis of observations usually leads to results that are affected by randomness. Therefore, it is helpful to compare results derived from data analyses with analytical models. Models represent ideas about distributions of random variables. Comparisons between measured and modeled probabilities may show whether such notions about the nature of randomness are applicable or not. In addition, results of data analyses are always related to a certain range, but it is often relevant to know the probability for the appearance of values of random variables outside the range of measured values. Such information requires the development of models for the distribution of random variables. The question is then on which basis it is possible to develop such models.

**Questions Considered.** The following sections in this chapter will address the questions regarding the analysis and modeling of random variables described in the preceding two paragraphs. There are two types of random variables: discrete and continuous variables. The difference between them is that discrete variables take on only certain values, whereas continuous variables may take on any values. For example, the results of tossing a coin (head or tail) or rolling the dice (numbers one to six) are discrete variables. Examples for continuous random variables are the human height and weight. The mathematical concepts for the analysis of discrete and continuous variables are very similar. In this and the following chapters we will only consider continuous random variables because they are the variables that are observed with regard to most problems in nature and technology.

## 4.2 Probability Density Functions

First, let us introduce and illustrate the basic concept used for the description of properties of random variables: a probability density function (PDF). This section provides the basis for the modeling and analysis concepts presented in Sects. 4.3 and 4.4.

### *4.2.1 Probability Density Functions*

**Mean Value.** We consider a random variable $X$, e.g., the intelligence quotient. There are $N$ measurement values $X_i$, where $i = 1, N$ (i.e., we know the intelligence quotients of $N$ people). A basic characterization of the random variable $X$ is given by the mean $\langle X \rangle$, which tells us which value we may expect for $X$. Therefore, the mean is often called an expectation value. To define $\langle X \rangle$ we use the definition applied in Chap. 2, this means the mean value $\langle X \rangle$ of $i = 1, N$ values of any variable $X_i$ is given by

$$\langle X \rangle = \frac{1}{N} \sum_{i=1}^{N} X_i \,. \tag{4.1}$$

A relevant question related to the definition (4.1) concerns the number of samples $N$ considered. The influence of $N$ on the calculation of means will be considered in detail in Sect. 4.4.2. Here, we assume that $N$ is sufficiently large.

**Moments and Central Moments.** In generalization of the definition of a mean value given by Eq. (4.1), we define a moment of $n^{\text{th}}$ order ($n = 1, 2, \ldots$) by

$$\langle X^n \rangle = \frac{1}{N} \sum_{i=1}^{N} X_i^n \,. \tag{4.2}$$

The setting $n = 1$ recovers Eq. (4.1), which means that a mean is the moment of first order. Usually, it is very helpful to analyze random variables by considering the mean and deviations from the mean

$$\widetilde{X}_i = X_i - \langle X \rangle, \tag{4.3}$$

which are called fluctuations. Such deviations from the mean can be characterized by the central moment (or the moment about the mean) of $n^{\text{th}}$ order,

$$\langle \widetilde{X}^n \rangle = \frac{1}{N} \sum_{i=1}^{N} \left( X_i - \langle X \rangle \right)^n \,. \tag{4.4}$$

A central moment of $n^{\text{th}}$ order can be expressed by the first $n$ moments by using the binomial theorem: see Eqs. (2.23) as an illustration. The definition (4.4) shows that the variance defined by Eq. (2.22b) is the central moment of second order. There is the obvious question of which information about random variables is given by the central moments of an order higher than 2. This question will be addressed below in conjunction with the discussion of several probability density functions.

**Distribution Function.** Information about the moments is usually very helpful, but such knowledge does not provide insight into the probability of finding certain events (e.g., the probability to have a person in a certain group of people with an

intelligence quotient higher than 180). Probabilities can be defined by means of theta functions, which are also called step functions or Heaviside functions. The theta function $\theta(z)$ of any variable $z$ can be defined by

$$\theta(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z \geq 0 \end{cases}. \tag{4.5}$$

We may replace $z$ by $z = x - X_i$. Here, $x$ represents any parameter, and $X_i$ is one measured value of a random variable. Then we obtain

$$\theta(x - X_i) = \begin{cases} 0 & \text{if } x - X_i < 0, \quad \text{i.e., if } x < X_i \\ 1 & \text{if } x - X_i \geq 0, \quad \text{i.e., if } X_i \leq x \end{cases}. \tag{4.6}$$

The probability for finding a value $X_i \leq x$ is one if $X_i \leq x$, and zero otherwise. This probability is reflected by the theta function $\theta(x - X_i)$. We have to take the mean value of $\theta(x - X_i)$ in order to find the probability $P(X \leq x)$ to find any $X \leq x$,

$$P(X \leq x) = \frac{1}{N}\sum_{i=1}^{N} \theta(x - X_i) = \langle \theta(x - X) \rangle. \tag{4.7}$$

To simplify the notation it is usual practice to introduce the distribution function (or cumulative distribution function)

$$F(x) = P(X \leq x) = \langle \theta(x - X) \rangle. \tag{4.8}$$

It is relevant to see here the difference between $X$ and $x$. $X$ is a random variable that is measurable, whereas the so-called sample space variable $-\infty \leq x \leq \infty$ is used to analyze the probability of $X$ values. Similar to Eq. (4.8), the probability $P(x \leq X \leq x + \Delta x)$ to find an $X$ value between $x$ and $x + \Delta x$ is given by

$$\begin{aligned} P(x \leq X \leq x + \Delta x) &= \langle \theta(x + \Delta x - X) - \theta(x - X) \rangle \\ &= P(X \leq x + \Delta x) - P(X \leq x) = F(x + \Delta x) - F(x). \end{aligned} \tag{4.9}$$

Here, $\Delta x$ is any positive interval that may be small or not.

**Distribution Function Properties.** The properties of the distribution function $F(x)$ can be seen in the following way. The probability $P(x \leq X \leq x + \Delta x)$ has to be non-negative according to its definition: for each data point $X_i$, the difference of theta functions $\theta(x + \Delta x - X_i) - \theta(x - X_i)$ can only be zero or one such that the mean of this difference has to be non-negative. Equation (4.9) then implies that

$$F(x) \leq F(x + \Delta x). \tag{4.10}$$

Thus, $F(x)$ increases with $x$. The minimum of the function $F(x) = <\theta(x - X)>$ is found for $F(-\infty)$. We find

$$F(-\infty) = 0 \tag{4.11}$$

because the minimum of the theta function is zero. The maximum of $F(x)$ is

$$F(\infty) = 1 \tag{4.12}$$

because the maximum of the theta function is one. Therefore, $F(x)$ is bounded by zero and one,

$$0 \le F(x) \le 1. \tag{4.13}$$

**Probability Density Function.** A probability density function (PDF) is defined as a derivative of the distribution function $F(x)$,

$$f(x) = \frac{dF(x)}{dx} = \left\langle \frac{d\theta(x - X)}{dx} \right\rangle. \tag{4.14}$$

Here, the definition $F(x) = <\theta(x - X)>$ is used to obtain the last expression. The meaning of the PDF is that $f(x)\,\mathrm{d}x$ determines the probability for finding $X$ in an infinitesimal interval between $x$ and $x + dx$,

$$P(x \le X \le x + dx) = F(x + dx) - F(x) = f(x)\,dx, \tag{4.15}$$

which follows from the Taylor expansion of $F(x + dx)$ for $dx \to 0$. Knowledge of $f(x)$ enables the calculation of all probabilities, for example

$$P(a \le X \le b) = \int_a^b f(x)\,dx = \int_a^b \frac{dF(x)}{dx}\,dx = F(b) - F(a). \tag{4.16}$$

The consistency with Eq. (4.9) can be seen by applying $x = a$ and $x + dx = b$ in Eq. (4.9). This formula can be used for the calculation of the distribution function $F(x)$ for any given $f(x)$. By replacing $a$ and $b$ by $-\infty$ and $x$, respectively, we find

$$F(x) = F(-\infty) + \int_{-\infty}^{x} f(y)\,dy = \int_{-\infty}^{x} f(y)\,dy, \tag{4.17}$$

where $F(-\infty) = 0$ is used.

**PDF Properties.** The PDF $f(x)$ has the following relevant properties,

$$f(x) \ge 0, \tag{4.18a}$$

$$f(-\infty) = f(\infty) = 0, \tag{4.18b}$$

$$\int f(x)\,dx = 1, \tag{4.18c}$$

$$\int g(x)\ f(x)\,dx = \langle g(X) \rangle. \tag{4.18d}$$

The first property is implied by the fact that $f(x)$ is the derivative of $F(x)$, which is a nondecreasing function of $x$. Thus, $f(x)$ cannot be negative. The second property results from the fact that $F(x)$ tends monotonically to zero or one as $|x| \to \infty$.

Hence, the derivative $f(x) = dF / dx$ tends to zero. The third property is a consequence of Eq. (4.16) for $a = -\infty$ and $b = \infty$. Because of $F(\infty) = 1$ and $F(-\infty) = 0$ we find $F(\infty) - F(-\infty) = 1$. The last property applies to any function $g(X)$. We may use, for example, $g = 1$, $g = X$, or $g = X^2$. For $g = 1$, Eq. (4.18d) recovers the relation (4.18c). The validity of Eq. (4.18d) for any function $g(X)$ can be seen by rewriting the left-hand side of Eq. (4.18d),

$$\begin{aligned}\int g(x)\, f(x)\, dx &= \int g(x)\, \frac{d\langle \theta(x - X)\rangle}{dx}\, dx = \frac{1}{N}\sum_{i=1}^{N} \int g(x)\, \frac{d\theta(x - X_i)}{dx}\, dx \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(-\frac{d}{dX_i}\right)\int g(x)\, \theta(x - X_i)\, dx \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(-\frac{d}{dX_i}\right)\int_{X_i}^{L} g(x)\, dx = \frac{1}{N}\sum_{i=1}^{N} g(X_i) = \langle g(X)\rangle.\end{aligned} \tag{4.19}$$

The first line makes use of the definition (4.14) of $f(x)$ and the definition of mean values. In the next line, the derivative by $x$ is replaced by the derivative by $-X_i$. Then, we reduce the integral to the range $x \geq X_i$ for which $\theta(x - X_i) \geq 0$, where $L \to \infty$. Finally, we make use of the Leibnitz integral rule to recover $<g(X)>$,

$$\frac{d}{dx}\int_{a(x)}^{b(x)} g(x,t)\, dt = \int_{a(x)}^{b(x)} \frac{\partial g(x,t)}{\partial x}\, dt + g\big(x, b(x)\big)\frac{db}{dx} - g\big(x, a(x)\big)\frac{da}{dx}. \tag{4.20}$$

Here, the derivatives are assumed to exist.

### *4.2.2 Delta Functions*

The theta function was used for the representation of the distribution function $F(x)$. The PDF $f(x) = <d\theta(x - X)/dx>$ was introduced as mean of the derivative of a theta function. Let us have a closer look at the properties of step functions and their derivatives, which will be helpful for the work with PDFs.

**Theta and Delta Functions as Limits.** Let us consider first the function $\theta_N(x)$ defined by the expression

$$\theta_N(x) = \frac{1 + \tanh(N x)}{2}. \tag{4.21}$$

Examples for $\theta_N(x)$ are shown in Fig. 4.2. This figure shows that $\theta_N(x)$ tends for large $N$ to the theta function

$$\theta(x) = \lim_{N\to\infty} \theta_N(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}. \tag{4.22}$$

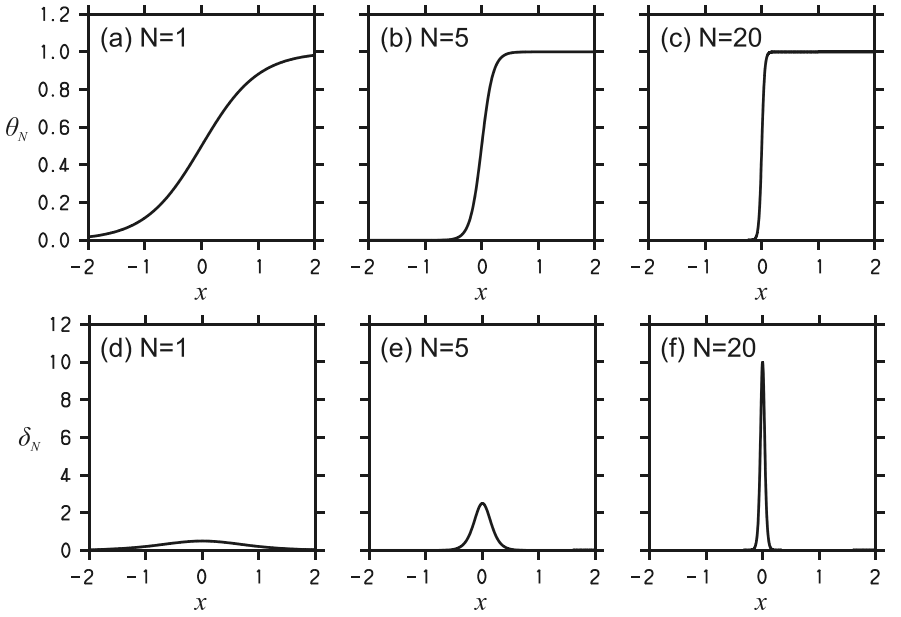



**Fig. 4.2.** The functions $\theta_N$ (first row) and $\delta_N$ (second row), which are defined by Eqs. (4.21) and (4.22), respectively, are shown for the given values of $N$.

To obtain an expression for $d\theta(x-X)/dx$ that we can use to see the properties of this derivative we introduce the function

$$\delta_N(x) = \frac{d\theta_N(x)}{dx}. \tag{4.23}$$

This relation does also apply to the limit that $N \to \infty$. In terms of Eq. (4.23) we obtain for the delta function $\delta(x)$ (or Dirac delta function or Dirac function)

$$\delta(x) = \lim_{N\to\infty} \delta_N(x) = \lim_{N\to\infty} \frac{d\theta_N(x)}{dx} = \frac{d}{dx} \lim_{N\to\infty} \theta_N(x) = \frac{d\theta(x)}{dx}. \tag{4.24}$$

The last expression arises from the definition (4.22) of the theta function as limit of $\theta_N(x)$ for $N \to \infty$. The advantage of introducing $\delta_N(x)$ is that we can obtain now a representation of $\delta(x) = d\theta(x-X)/dx$ in terms of the limit

$$\delta(x) = \lim_{N\to\infty} \delta_N(x) = \lim_{N\to\infty} \frac{d\theta_N(x)}{dx} = \lim_{N\to\infty} \frac{N}{2\cosh^2(Nx)}. \tag{4.25}$$

The last expression can be found by taking the derivative of $\theta_N(x)$ defined by Eq. (4.21). Examples for $\delta_N(x)$ are given in Fig. 4.2. This figure shows that $\delta(x)$ vanishes for all $x \neq 0$, and it diverges for $x \to 0$. For that reason, the delta function is referred to as a generalized function or a distribution.

**Properties.** A generalized function is a function that does not have to exist for all arguments, but integrals over such functions (multiplied with other functions) have to exist. The delta function has the property that it integrates to one,

$$\int \delta(x)\,dx = \int \frac{d\theta}{dx}\,dx = \theta(\infty) - \theta(-\infty) = 1. \tag{4.26}$$

The most useful property of the delta function is its so-called sifting property,

$$\int g(x)\,\delta(x-a)dx = g(a)\int \delta(x-a)dx = g(a). \tag{4.27}$$

Here, $g(x)$ can be any function of $x$, and $a$ is any finite real number. Justification for the middle expression arises from the fact $\delta(x - a)$ vanishes everywhere except at $x = a$. Therefore, $g(x)$ can be replaced by $g(a)$ that is independent of $x$. The last expression results from the normalization property given by Eq. (4.26). Another relevant property of the delta function is given by

$$\int g(x)\,\delta(-x)dx = \int g(x)\,\delta(x)dx. \tag{4.28}$$

The right-hand side is equal to the left-hand side because the only nonzero contribution of the delta function is at $x = 0$. There are many other properties of delta functions – useful additional information about delta functions can be found elsewhere (Heinz 2003).

**PDF Definition.** According to the PDF definition $f(x) = \langle d\theta(x - X)/dx \rangle$, the expression $\delta(x) = d\theta(x - X)/dx$ can be used for the following definition of a PDF,

$$f(x) = \big\langle \delta(x - X) \big\rangle. \tag{4.29}$$

This definition can be used to recover all the PDF properties. Property (4.18a) is a consequence of the fact that $f(x)$ is the mean of a non-negative variable. Property (4.18b) is caused by the fact that the delta function vanishes at infinity,

$$f(\pm\infty) = \big\langle \delta(\pm\infty - X) \big\rangle = 0. \tag{4.30}$$

Property (4.18c) is implied by the normalization of delta functions, which is given by Eq. (4.26),

$$\int f(x)\,dx = \int \big\langle \delta(x - X) \big\rangle\,dx = \left\langle \int \delta(x - X)\,dx \right\rangle = \big\langle 1 \big\rangle = 1. \tag{4.31}$$

Property (4.18d) is implied by the sifting property (4.27) of the delta function,

$$\begin{aligned} \int g(x)\, f(x)\,dx &= \int g(x)\, \big\langle \delta(x - X) \big\rangle\,dx = \left\langle \int g(x)\, \delta(x - X)\,dx \right\rangle \\ &= \left\langle \int g(X)\, \delta(x - X)\,dx \right\rangle = \left\langle g(X) \int \delta(x - X)\,dx \right\rangle = \big\langle g(X) \big\rangle. \end{aligned} \tag{4.32}$$

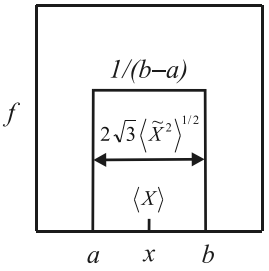



**Fig. 4.3.** An illustration of a uniform PDF that has a constant value $1/(b-a)$ between $a$ and $b$. The mean value $<X> = (a+b)/2$, and the standard deviation $<\tilde{X}^2>^{1/2} = (b-a)/(2\,3^{1/2})$.

### *4.2.3 An Example: The Uniform Probability Density Function*

**Uniform PDF.** It is relatively often the case that there is no information about the distribution of a random variable. In this case, it is reasonable to assume that there is a uniform probability for all values of this random variable inside a certain range of variations. The model for such a PDF is given by a uniform PDF,

$$f(x) = \begin{cases} 1/(b-a) & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}. \tag{4.33}$$

Here, the parameters $a$ and $b$ define the range of nonzero probability. This PDF satisfies the normalization condition to integrate to unity,

$$\int f(x)\,dx = \frac{1}{b-a}\int_a^b dx = 1. \tag{4.34}$$

**Probability.** To calculate the probability to find $X$ between any two bounds we calculate first the distribution function $F(x)$. By means of Eq. (4.17) we find

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \int_{-\infty}^{x} \frac{\theta(b-y)-\theta(a-y)}{b-a}\,dy = \frac{\min(x,b)-\min(x,a)}{b-a}. \tag{4.35}$$

Here, the PDF is represented in terms of a difference of theta functions that is only nonzero if $a \le y \le b$. The use of $P(c \le X \le d) = F(d) - F(c)$ then results in the following expression for the probability to find $X$ between $c$ and $d$,

$$P(c \le X \le d) = F(d) - F(c) = \frac{\min(d,b)-\min(d,a)-\min(c,b)+\min(c,a)}{b-a}. \tag{4.36}$$

Two examples for the use of this expression are the following ones:

$$P(c \le X \le d) = \left\{ \begin{array}{ll} \dfrac{b-a-c+c}{b-a} = 1 & \text{if } c < a < b < d \\ \dfrac{d-a-c+a}{b-a} = \dfrac{d-c}{b-a} & \text{if } a < c < d < b \end{array} \right\}. \tag{4.37}$$

**Moments.** The mean value of a uniform PDF $f(x)$ is given by

$$\langle X\rangle = \int x\, f(x)\, dx = \frac{1}{b-a}\int_a^b x\, dx = \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{1}{2}\frac{(b-a)(b+a)}{b-a} = \frac{a+b}{2}. \tag{4.38}$$

Therefore, $<X>$ is the mean position between $a$ and $b$, see Fig. 4.3. The central moments can be calculated in the following way ($k = 1, 2, \ldots$),

$$\begin{aligned}\left\langle \widetilde{X}^k \right\rangle &= \int \left(x - \langle X\rangle\right)^k f(x)\, dx = \frac{1}{b-a}\int_a^b \left(x - \frac{a+b}{2}\right)^k dx \\ &= \frac{1}{(k+1)(b-a)}\left[\left(\frac{b-a}{2}\right)^{k+1} - \left(\frac{a-b}{2}\right)^{k+1}\right] \\ &= \frac{1}{2(k+1)}\frac{2}{(b-a)}\left[\left(\frac{b-a}{2}\right)^{k+1} - (-1)^{k+1}\left(\frac{b-a}{2}\right)^{k+1}\right] = \frac{1-(-1)^{k+1}}{2(k+1)}\left(\frac{b-a}{2}\right)^k.\end{aligned} \tag{4.39}$$

Hence, all the odd central moments disappear ($k = 1, 2, \ldots$),

$$\left\langle \widetilde{X}^{2k-1} \right\rangle = 0. \tag{4.40}$$

The even moments are determined by the formula ($k = 1, 2, ...$)

$$\left\langle \widetilde{X}^{2k} \right\rangle = \frac{1}{2k+1}\left(\frac{b-a}{2}\right)^{2k} = \frac{3^k}{2k+1}\left\langle \widetilde{X}^2 \right\rangle^k. \tag{4.41}$$

The last expression was obtained by adopting the second central moment

$$\left\langle \widetilde{X}^2 \right\rangle = \frac{1}{3}\left(\frac{b-a}{2}\right)^2. \tag{4.42}$$

According to this expression, the standard deviation $< \widetilde{X}^2 >^{1/2}$ is given by

$$\left\langle \widetilde{X}^2 \right\rangle^{1/2} = \frac{b-a}{2\sqrt{3}}. \tag{4.43}$$

Here, we assumed $b \geq a$. $< \widetilde{X}^2 >^{1/2}$ characterizes the typical amount of deviations from the mean. The standard deviation scales the width of the PDF, but there is no need that $< \widetilde{X}^2 >^{1/2}$ is equal to the PDF width (see Fig. 4.3).

**Parameters.** For given values of the mean and standard deviation we can calculate the model parameters $a$ and $b$. According to Eqs. (4.38) and (4.43) we have

$$b + a = 2\langle X\rangle, \qquad b - a = 2\sqrt{3}\left\langle \widetilde{X}^2 \right\rangle^{1/2}. \tag{4.44}$$

Hence, $a$ and $b$ are determined by the mean and standard deviation according to

$$b = \langle X\rangle + \sqrt{3}\left\langle \widetilde{X}^2 \right\rangle^{1/2}, \qquad a = \langle X\rangle - \sqrt{3}\left\langle \widetilde{X}^2 \right\rangle^{1/2}, \tag{4.45}$$

as may be seen by taking the sum and difference of $a$ and $b$.

**Random Number Generation.** The relevance of the uniform PDF also arises from the fact that random numbers having a uniform PDF, which can be generated in several ways (L'Ecuyer 1994), can be used to generate random numbers having different continuous distributions. To show this, we consider a random variable $X$ that is uniformly distributed over the interval [0, 1], i.e., the PDF of $X$ is given by

$$f(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}. \tag{4.46}$$

The corresponding distribution function of $X$ is given by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(z)\,dz = \int_{-\infty}^{x} \left[\theta(1-z) - \theta(0-z)\right] dz = \min(x,1) - \min(x,0). \tag{4.47}$$

Here, the PDF $f(x)$ is written in terms of the bracket term, which is only nonzero for $0 \le z \le 1$. The latter result can also be written

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x \le 1. \\ 1 & \text{if } x > 1 \end{cases} \tag{4.48}$$

We are interested to generate random numbers that have a specified distribution function $G(y)$. Then, it turns out that the random variable

$$Y = G^{-1}(X), \tag{4.49}$$

which can be determined by solving the relation $G(Y) = X$ for $Y$, has the given distribution function $G(y)$. The correctness of this claim can be shown in the following way. The distribution function of $Y = G^{-1}(X)$ is given by

$$F(y) = P(Y \le y) = P\left(G^{-1}(X) \le y\right) = P\left(X \le G(y)\right). \tag{4.50}$$

The last expression arises from the fact that $F(y)$ is a monotonic function. Thus, the inequality $G^{-1}(X) \le y$ is satisfied if and only if $X \le G(y)$. $P(X \le G(y))$ can be calculated by means of the known distribution function of $X$. By replacing $x$ in Eq. (4.48) by $G(y)$ we obtain

$$F(y) = P(X \le G(y)) = \begin{cases} 0 & \text{if } G(y) < 0 \\ G(y) & \text{if } 0 \le G(y) \le 1. \\ 1 & \text{if } G(y) > 1 \end{cases} \tag{4.51}$$

Due to the fact that $0 \le G(y) \le 1$, this relation implies that

$$F(y) = G(y). \tag{4.52}$$

Therefore, the random variable $Y$ has the given distribution function $G(y)$. This method is called the inverse transformation method (Ross 2010).

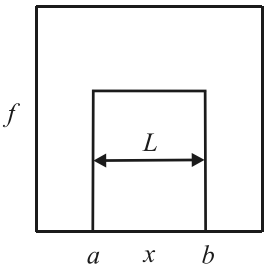



**Fig. 4.4.** The PDF $f(x)$ of a uniformly distributed random variable $x$. Here, $L = b - a$ refers to the interval in which the PDF is nonzero.

## 4.3 Models for Probability Density Functions

After defining PDFs and discussing their basic properties, let us consider the question of how it is possible to find specific PDF shapes for certain observations. There are two main approaches to address this question: we can design PDFs on the basis of any principle (e.g., the constraint considered below that the PDF has to maximize the uncertainty), or we apply empirical PDF shapes that have desired properties. The first approach will be explained in Sects. 4.3.1 and 4.3.2, which includes the discussion of the most commonly used PDF – the normal PDF. This approach will be presented for unbounded variables that may have values between negative and positive infinity. The second approach will be used in Sects. 4.3.3 and 4.3.4 to obtain PDFs for the relevant cases that the variables can only be positive and that the variables can only vary between zero and one, respectively.

### *4.3.1 Statistically Most-Likely Probability Density Functions*

**Predictability.** For the development of PDF models it is helpful to relate the shape of a PDF to a measure that characterizes the predictability of the state of a random variable. The predictability of states is illustrated in Fig. 4.4, which shows that the predictability of the state of a random variable depends on the PDF shape, which is controlled by $L = b - a$. For $L \to 0$, the PDF becomes a delta function. The predictability is maximal in this case (the uncertainty is minimal): we know that $X$ will realize the value at the location of the delta peak. For $L \to \infty$, the predictability of the state of $X$ is minimal (the uncertainty is maximal): there is the same probability for all possible states. The consideration of the predictability of states of random variables can be used in the following way for the construction of PDFs. Very often we do not know anything about a PDF with the exception of a

few low-order moments (means and variances), for which reasonable estimates are often available (contrary to measurements of moments of higher order that are usually significantly affected by randomness). However, knowledge of a few low-order moments is insufficient to determine the PDF shape, which requires information about all the moments. What we can do in this case is to apply information about the known moments combined with the constraint that the predictability (uncertainty) related to the PDF has to be minimal (maximal). The latter approach corresponds to the following idea (Du et al. 1994). First, we reduce our uncertainty by the given information (the known moments). Second, we are maximally uncommitted with respect to the missing information (the PDF shape). The use of this concept will be shown in the following. Interestingly, this approach can be extended to develop PDF models for several variables, and it can be also used for bounded variables (see the exercises 4.3.5 – 4.3.8).

**Measure of Uncertainty.** The realization of this concept requires the definition of a measure $S$ of the uncertainty of the state of a random variable. This measure of uncertainty $S$, which is called entropy, is defined by

$$S = -\int f(x) \ln\big(f(x)\big) dx . \tag{4.53}$$

Instead of justifying this definition of $S$ in detail (Shannon 1948, Jaynes 1957), let us illustrate the suitability of this definition with regard to a uniform PDF. The combination of the definition of $S$ with the uniform PDF shape (4.33) shows that the entropy of a uniform PDF is given by the expression

$$S = -\frac{1}{b-a}\int_a^b \ln\left(\frac{1}{b-a}\right) dx = \frac{\ln(b-a)}{b-a}(b-a) = \ln L , \tag{4.54}$$

where $L = b - a$ is used. This expression illustrates the suitability of using $S$ as a measure of uncertainty: the uncertainty $S$ is minimal for $L \to 0$, and the uncertainty is maximal for $L \to \infty$.

**Statistically Most-Likely PDF.** We will construct now a PDF according to the concept described above. We assume that we know $n = 1, \cdots, s$ moments

$$\left\langle X^n \right\rangle = \int x^n f(x) dx . \tag{4.55}$$

The goal is to construct a PDF that has $s$ moments that agree with the given ones but maximizes the entropy $S$ (i.e., the uncertainty). According to the calculus of variations, we extend the entropy $S$ to a functional $S^*$ by involving Eq. (4.55),

$$S^* = -\int \left\{ f(x) \ln\big(f(x)\big) + \sum_{k=0}^{s} \mu_k x^k \, f(x) - f(x) \right\} dx . \tag{4.56}$$

The $\mu_k$ are Lagrange multipliers which have to be chosen such that Eq. (4.55) is satisfied for all $n$. The last term on the right-hand side of Eq. (4.56) modifies the multiplier $\mu_0$, which simplifies the following explanations. To find the maximum

of $S^*$, we calculate the functional variation of $S^*$ with regard to $f$,

$$\frac{\delta S^*}{\delta f} = -\int \left\{ \ln\big(f(x)\big) + \sum_{k=0}^{s} \mu_k x^k \right\} dx. \tag{4.57}$$

This functional variation is equal to zero if the PDF $f(x)$ is given by

$$f(x) = \exp\left\{ -\sum_{k=0}^{s} \mu_k x^k \right\}. \tag{4.58}$$

The PDF (4.58) is called a statistically most-likely (SML) PDF. By introducing nondimensional Lagrange multipliers $\lambda_k$ we can also write the latter relation as

$$f(x) = \exp\left\{ \lambda_0 - \sum_{k=1}^{s} \frac{\lambda_k}{k} \frac{\big(x - \langle X \rangle\big)^k}{\big\langle \widetilde{X}^2 \big\rangle^{k/2}} \right\}. \tag{4.59}$$

The $s + 1$ factors $\lambda_k$ are uniquely determined by the normalization of $f(x)$ and the $s$ conditions (4.55). These conditions for the PDF (4.59) can be written

$$\big\langle \widetilde{X}^n \big\rangle = \int \big(x - \langle X \rangle\big)^n f(x)\, dx, \tag{4.60}$$

where $n = 0, 1, \cdots, s$. Evidence that the PDF (4.59) maximizes the entropy is provided in terms of the exercise 4.3.1 by the comparison of the entropy of a SML PDF with the entropy of the corresponding uniform PDF.

### *4.3.2 The Normal Probability Density Function*

**Second-Order SML PDF.** The normal PDF $f(x)$ of one random variable $X$ is a second-order SML PDF. This PDF is given by

$$f(x) = \exp\left\{ \lambda_0 - \lambda_1 \frac{x - \langle X \rangle}{\big\langle \widetilde{X}^2 \big\rangle^{1/2}} - \frac{\lambda_2}{2} \frac{\big(x - \langle X \rangle\big)^2}{\big\langle \widetilde{X}^2 \big\rangle} \right\}. \tag{4.61}$$

The parameters $\lambda_0$, $\lambda_1$, and $\lambda_2$ have to be chosen such that the PDF (4.61) satisfies Eq. (4.60) for $n = 0, 1, 2$,

$$1 = \int f(x)\, dx, \tag{4.62a}$$

$$0 = \int \big(x - \langle X \rangle\big) f(x)\, dx, \tag{4.62b}$$

$$\big\langle \widetilde{X}^2 \big\rangle = \int \big(x - \langle X \rangle\big)^2 f(x)\, dx, \tag{4.62c}$$

for any given mean $<X>$ and variance $< \widetilde{X}^2 >$.

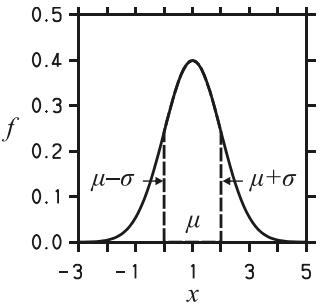



**Fig. 4.5.** The normal PDF $f(x)$ given by Eq. (4.72) with mean $\mu = 1$ and standard deviation $\sigma = 1$.

**Parameter Calculation.** A simple way to calculate the model parameters is to differentiate $f(x)$,

$$\frac{df}{dx} = \frac{f}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}} \left\{ -\lambda_1 - \lambda_2 \frac{x - \langle X \rangle}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}} \right\}. \tag{4.63}$$

The integration of this relation leads to

$$\int \frac{df}{dx} dx = -\frac{1}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}} \int \left\{ \lambda_1 + \lambda_2 \frac{x - \langle X \rangle}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}} \right\} f \, dx = -\frac{\lambda_1}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}}, \tag{4.64}$$

where Eqs. (4.62a) and (4.62b) are applied. The left-hand side disappears because the PDF $f(x)$ disappears at infinity. Therefore, we find the requirement $\lambda_1 = 0$. By multiplying Eq. (4.63) for $\lambda_1 = 0$ with $x - \langle X \rangle$ and integration we find

$$\int \left( x - \langle X \rangle \right) \frac{df}{dx} dx = -\frac{\lambda_2}{\left\langle \widetilde{X}^2 \right\rangle} \int \left( x - \langle X \rangle \right)^2 f \, dx = -\lambda_2, \tag{4.65}$$

where Eq. (4.62c) is used. The left-hand side can be written

$$\begin{aligned} \int \left( x - \langle X \rangle \right) \frac{df}{dx} dx &= \int \left[ \frac{d\left( x - \langle X \rangle \right) f}{dx} - \frac{d\left( x - \langle X \rangle \right)}{dx} f \right] dx \\ &= \left[ \left( x - \langle X \rangle \right) f \right]_{\infty} - \left[ \left( x - \langle X \rangle \right) f \right]_{-\infty} - \int f \, dx = -1. \end{aligned} \tag{4.66}$$

The term $(x - \langle X \rangle) f$ is zero at positive and negative infinity, see Eq. (4.61). The combination of the last two relations leads then to the conclusion that $\lambda_2 = 1$. The parameter $\lambda_0$ can be calculated by means of the condition (4.62a),

$$1 = \mathrm{e}^{\lambda_0} \int \exp\left\{ -\frac{1}{2} \frac{\left( x - \langle X \rangle \right)^2}{\left\langle \widetilde{X}^2 \right\rangle} \right\} dx, \tag{4.67}$$

where the findings $\lambda_1 = 0$ and $\lambda_2 = 1$ are applied. Now, we introduce the variable $y = (x - <X>) / (2 < \widetilde{X}^2 >)^{1/2}$ to simplify the latter relation,

$$e^{-\lambda_0} = \sqrt{2\left\langle \widetilde{X}^2 \right\rangle} \int e^{-y^2} dy. \tag{4.68}$$

To calculate the integral involved it is helpful to consider the squared integral

$$\left[\int e^{-y^2} dy\right]^2 = \int e^{-y^2} dy \int e^{-z^2} dz = \int\int e^{-(y^2+z^2)} dy\, dz. \tag{4.69}$$

The right-hand side integral can be calculated by introducing the polar coordinates $y = \mathrm{r} \cos\theta$ and $z = \mathrm{r} \sin\theta$. In terms of $dy\, dz = r\, d\theta\, dr$ we obtain

$$\left[\int e^{-y^2} dy\right]^2 = \int_0^\infty \int_0^{2\pi} e^{-r^2} r\, d\theta\, dr = 2\pi \int_0^\infty r e^{-r^2} dr = -\pi \int_0^\infty \frac{de^{-r^2}}{dr} dr = \pi. \tag{4.70}$$

The combination of this result with Eq. (4.68) then shows that

$$e^{-\lambda_0} = \sqrt{2\pi\left\langle \widetilde{X}^2 \right\rangle}. \tag{4.71}$$

The second-order SML PDF (4.61) that follows from the latter result and $\lambda_1 = 0$ and $\lambda_2 = 1$ can be written in terms of standard notation as

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}. \tag{4.72}$$

Here, the model parameters $\mu$ and $\sigma$ are given by

$$\mu = \left\langle X \right\rangle, \qquad \sigma = \left\langle \widetilde{X}^2 \right\rangle^{1/2}. \tag{4.73}$$

The second-order SML PDF (4.72) is the normal PDF. The PDF $f(x)$ is illustrated in Fig. 4.5. This figure shows the relevance of the mean: $<X> = \mu$ describes the expectation value, which is here the peak value of the PDF. The square root of the variance $\sigma = < \widetilde{X}^2 >^{1/2}$ is the standard deviation. Figure 4.5 shows the meaning of the standard deviation: this parameter characterizes the intensity of fluctuations (the amount of randomness). In correspondence to the properties of the uniform PDF we observe that $\sigma$ does not determine all the range of fluctuations, but it defines the order of magnitude of fluctuations. The PDF is a symmetric function about the mean. Thus, all the odd central moments disappear ($k = 1, 2, ...$),

$$\left\langle \widetilde{X}^{2k-1} \right\rangle = 0. \tag{4.74}$$

The even central moments are determined by the formula ($k = 1, 2, ...$)

$$\left\langle \widetilde{X}^{2k} \right\rangle = \frac{(2k)!}{2^k\, k!} \sigma^{2k}. \tag{4.75}$$

**Skewness and Flatness.** To compare any PDF with a normal PDF, or to see whether a random data set can be described by a normal PDF, it is helpful to use normalized moments of third-order (the skewness) and fourth-order (the flatness or kurtosis) as a reference. These nondimensional numbers are defined by

$$m_3 = \left\langle \widetilde{X}^3 \right\rangle \left\langle \widetilde{X}^2 \right\rangle^{-3/2}, \qquad m_4 = \left\langle \widetilde{X}^4 \right\rangle \left\langle \widetilde{X}^2 \right\rangle^{-2}. \tag{4.76}$$

The skewness $m_3$ indicates deviations from the symmetry of fluctuations about the mean value. According to Eq. (4.74), one finds for the normal PDF

$$m_3 = 0. \tag{4.77}$$

The flatness $m_4$ indicates the peakedness of a PDF. According to Eq. (4.75), a normal PDF is characterized by

$$m_4 = 3. \tag{4.78}$$

A PDF with $m_4 \geq 3$ ($m_4 \leq 3$) has a higher (lower) peak value than the normal PDF. It is worth noting that $m_4$ and $m_3$ cannot take any values: they have to satisfy the inequality $m_3^2 + 1 \leq m_4$ (see Eq. (2.38)).

**Probability.** The probability $P(a \leq X \leq b)$ for finding the random variable $X$ between $a$ and $b$ is defined by Eq. (4.16). For a normal PDF we find

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx = \frac{1}{\sqrt{2\pi}\,\sigma} \int_a^b \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} dx. \tag{4.79}$$

We introduce $y = (x-\mu)/(2^{1/2}\,\sigma)$ to simplify this expression,

$$P(a \leq X \leq b) = \frac{1}{\sqrt{\pi}} \int_A^B e^{-y^2}\,dy, \tag{4.80}$$

where the bounds are given by $A = (a-\mu)/(2^{1/2}\,\sigma)$ and $B = (b-\mu)/(2^{1/2}\,\sigma)$. The integral (4.80) cannot be explicitly calculated. The most convenient way to deal with this problem is to write $P(a \leq x \leq b)$ in terms of a standardized integral. To prepare this representation we write $P(a \leq x \leq b)$ as

$$P(a \leq X \leq b) = \frac{1}{\sqrt{\pi}} \left[ \int_A^0 e^{-y^2}\,dy + \int_0^B e^{-y^2}\,dy \right] = \frac{1}{\sqrt{\pi}} \left[ \int_0^B e^{-y^2}\,dy - \int_0^A e^{-y^2}\,dy \right]. \tag{4.81}$$

The latter expression can be written

$$P(a \leq X \leq b) = \frac{1}{2} \left[ erf(B) - erf(A) \right], \tag{4.82}$$

where the error function $erf(x)$ is defined by the expression

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2}\,ds. \tag{4.83}$$

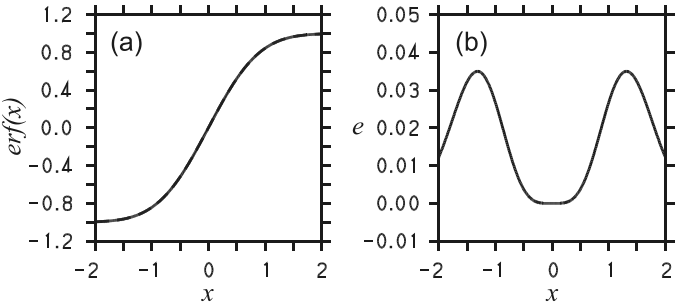



**Fig. 4.6.** (a) The error function *erf*(*x*) defined by Eq. (4.83) is given as a *solid line*, the approximation (4.84) as a *dashed line* (no noticeable difference between these curves); (b) the relative error *e* (in %) of the approximation (4.84).

**Error Function Approximation.** The error function *erf*(*x*) can be numerically calculated in terms of its integral definition (4.83). However, the use of an analytical approximation is much more convenient. Such an approximation is given by

$$E(x) = \pm\sqrt{1 - e^{-Hx^2}}\,. \tag{4.84}$$

Here, the abbreviation $H$ is defined by

$$H = \frac{4/\pi + p\,x^2}{1 + p\,x^2}, \tag{4.85}$$

and the parameter $p$ is given by

$$p = -\frac{8}{3\pi}\frac{\pi - 3}{\pi - 4} = 0.1400. \tag{4.86}$$

The positive sign in Eq. (4.84) applies to positive $x$ values, and the negative sign applies to negative $x$ values. Figure 4.6 shows that the approximation (4.84) is very good: the magnitude of the relative error of this approximation is smaller than 0.04%; the figure also reveals the relevant property $erf(-x) = -erf(x)$ of the error function and of the approximation $E(x)$ of the error function.

**Example Probabilities.** Let us calculate some probabilities to illustrate the use of Eq. (4.82) combined with the approximation (4.84) for the error function,

$$P(\mu - \sigma \le X \le \mu + \sigma) = \frac{erf(1/2^{1/2}) - erf(-1/2^{1/2})}{2} = erf(2^{-1/2}) = 0.683,$$

$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = \frac{erf(2/2^{1/2}) - erf(-2/2^{1/2})}{2} = erf(2/2^{1/2}) = 0.955,$$

$$P(\mu - 3\sigma \le X \le \mu + 3\sigma) = \frac{erf(3/2^{1/2}) - erf(-3/2^{1/2})}{2} = erf(3/2^{1/2}) = 0.997. \tag{4.867}$$

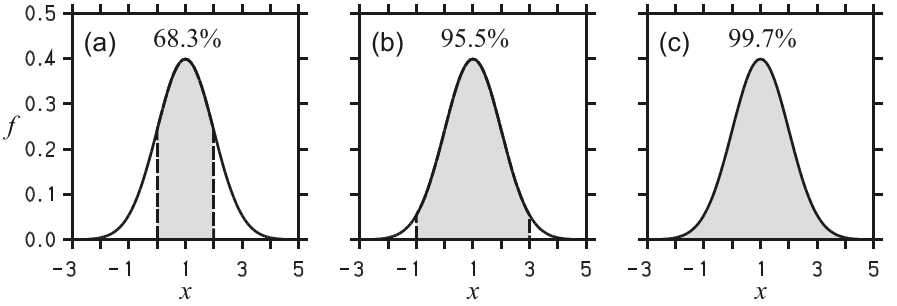



**Fig. 4.7.** A normal PDF $f(x)$ with mean $\mu = 1$ and standard deviation $\sigma = 1$. (a), (b), and (c) illustrate the probabilities $P(\mu - \sigma \le X \le \mu + \sigma) = 68.3\%$, $P(\mu - 2\sigma \le X \le \mu + 2\sigma) = 95.5\%$, and $P(\mu - 3\sigma \le X \le \mu + 3\sigma) = 99.7\%$, respectively.

$A$ and $B$ are determined by $A = (a - \mu) / (2^{1/2}\, \sigma)$ and $B = (b - \mu) / (2^{1/2}\, \sigma)$, and the property $erf(-x) = -erf(x)$ was used here to simplify these expressions. The corresponding probabilities are calculated in terms of the approximation $E(x)$ of the error function. Figure 4.7 illustrates the relevance of these probabilities. The probability for finding $X$ between $\mu - 3\sigma$ and $\mu + 3\sigma$ is close to one. Consequently, the magnitude of almost all fluctuations will be smaller than $3\sigma$. A mean value that characterizes both small and large fluctuations is given by the standard deviation $< \widetilde{X}^2 >^{1/2} = \sigma$. Thus, $\sigma$ is used as a measure for the typical intensity of fluctuations. The probability to find $X$ between $\mu - \sigma$ and $\mu + \sigma$ is about 2 / 3.

**Application of the Normal PDF.** The normal distribution is used for a variety of applications, for example the distribution of

- intelligence quotients, where $\mu = 100$, $\sigma = 15$ (Bulmer 1979, Stewart 2006)
- heights of adult males in the U.S., where $\mu = 1.75$ m, $\sigma = 0.07$ m (Stewart 2006)
- lengths of human pregnancies, where $\mu = 268$ days, $\sigma = 15$ days (Stewart 2006)
- test scores, where, e.g., $\mu = 80\%$, $\sigma = 10\%$

Other examples for normal PDFs will be shown below.

### *4.3.3 The Gamma Probability Density Function*

There are many circumstances under which the application of a normal PDF is inappropriate. One example is given by the relevant case that the random variable is non-negative. The latter case is given, e.g., regarding the atmospheric raindrop size distribution and the droplet size distribution in a cloud (both distributions are relevant to climate modeling). A way to handle such cases will be considered now by the following discussion of the gamma PDF.

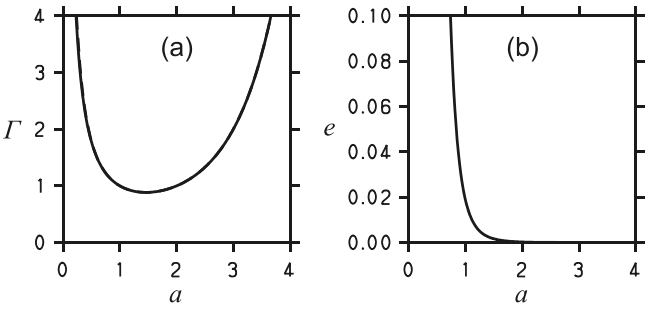



**Fig. 4.8.** The *solid* and *dashed lines* in (a) show the gamma function (4.89) and its approximation (4.90), respectively. There is hardly any difference between both curves; (b) shows the relative error $e$ (in %) of the approximation (4.90).

**Gamma PDF.** The PDF of non-negative random variables is often modeled by the gamma PDF. This PDF is defined by

$$f(x) = \begin{cases} \dfrac{b}{\Gamma(a)\langle X\rangle}\left(\dfrac{b\,x}{\langle X\rangle}\right)^{a-1}\exp\left(-\dfrac{b\,x}{\langle X\rangle}\right) & \text{if } 0 \le x \\ 0 & \text{otherwise} \end{cases} . \tag{4.88}$$

The idea of this PDF is to consider an exponential function $\exp(-b\,x/\langle X\rangle)$ that is multiplied with a power of $b\,x/\langle X\rangle$ to increase the flexibility of PDF shapes (to cover more scenarios that can be described). The PDF parameters $a$ and $b$ are non-dimensional. $\Gamma(a)$ is the gamma function that is defined by the integral

$$\Gamma(a) = \int_0^\infty y^{a-1}e^{-y}dy. \tag{4.89}$$

This integral cannot be solved analytically, but it can be approximated by

$$\Gamma(a) = \sqrt{\frac{2\pi}{a}}\left(\frac{a}{e}\sqrt{a\sinh\left(\frac{1}{a}\right)+\frac{1}{810\,a^6}}\right)^a . \tag{4.90}$$

Figure 4.8 shows that Eq. (4.90) approximates the gamma function very well. For example, the relative error at $a = (1, 2)$ is $e = (0.0183, 0.0003)\%$, respectively. The gamma function has the relevant property

$$\Gamma(a+1) = a\,\Gamma(a), \tag{4.91}$$

which can be used to calculate the gamma function at higher $a$ – the accuracy of (4.90) increases with $a$. The latter property can be proven by integration by parts,

$$\int_0^\infty \frac{dy^a e^{-y}}{dy}dy = 0 = a\int_0^\infty y^{a-1}e^{-y}dy - \int_0^\infty y^a e^{-y}dy. \tag{4.92}$$

The reason for the appearance of the gamma function in Eq. (4.88) can be seen by proving that the gamma PDF satisfies the normalization condition,

$$\int_0^\infty f(x)\,dx = \frac{b}{\Gamma(a)\langle X\rangle}\int_0^\infty\left(\frac{b\,x}{\langle X\rangle}\right)^{a-1}\exp\left(-\frac{b\,x}{\langle X\rangle}\right)dx = \frac{1}{\Gamma(a)}\int_0^\infty y^{a-1}e^{-y}\,dy = 1, \quad (4.93)$$

where the substitution $y = b\,x/\langle X\rangle$ was applied.

**Moments.** The moments of $f(x)$ can be calculated in this way ($k = 1, 2, \ldots$),

$$\begin{aligned}\left\langle X^k\right\rangle &= \int_0^\infty x^k\,f(x)\,dx = \frac{b}{\Gamma(a)\langle X\rangle}\frac{\langle X\rangle^k}{b^k}\int_0^\infty\left(\frac{b\,x}{\langle X\rangle}\right)^k\left(\frac{b\,x}{\langle X\rangle}\right)^{a-1}\exp\left(-\frac{b\,x}{\langle X\rangle}\right)dx \\ &= \frac{1}{\Gamma(a)}\frac{\langle X\rangle^k}{b^k}\int_0^\infty y^{a+k-1}e^{-y}\,dy = \frac{\langle X\rangle^k}{b^k}\frac{\Gamma(a+k)}{\Gamma(a)}.\end{aligned} \quad (4.94)$$

Here, $y = b\,x/\langle X\rangle$ and the definition (4.89) of the gamma function were applied. The ratio of gamma functions, which is often denoted by Pochmammer's symbol $(a)_k = \Gamma(a+k)/\Gamma(a)$, can be rewritten by making use of $\Gamma(a+1) = a\,\Gamma(a)$,

$$\begin{aligned}\frac{\Gamma(a+k)}{\Gamma(a)} &= \frac{(a+k-1)\Gamma(a+k-1)}{\Gamma(a)} = \frac{(a+k-1)(a+k-2)\Gamma(a+k-2)}{\Gamma(a)} \\ &= (a+k-1)(a+k-2)\cdots a.\end{aligned} \quad (4.95)$$

Hence, the moments of $f(x)$ are given by

$$\left\langle X^k\right\rangle = \frac{(a+k-1)(a+k-2)\cdots a}{b^k}\langle X\rangle^k. \quad (4.96)$$

**Parameters.** The last expression shows that the moments of $f(x)$ are functions of the parameters $a$ and $b$. These relations can be used for the calculation of $a$ and $b$ as functions of the mean and the variance, which enables the adjustment of the gamma PDF to any given values for the mean and the variance. According to Eq. (4.96), the mean is given by

$$\langle X\rangle = \frac{a}{b}\langle X\rangle, \quad (4.97)$$

and the second-order moment is

$$\left\langle X^2\right\rangle = \frac{(a+1)\,a}{b^2}\langle X\rangle^2. \quad (4.98)$$

The variance $\langle\widetilde{X}^2\rangle$ is related to these moments by $\langle\widetilde{X}^2\rangle = \langle X^2\rangle - \langle X\rangle^2$, see Eq. (2.23a). Applying the last two relations, we obtain for the variance

$$\left\langle\widetilde{X}^2\right\rangle = \frac{a}{b^2}\langle X\rangle^2. \quad (4.99)$$

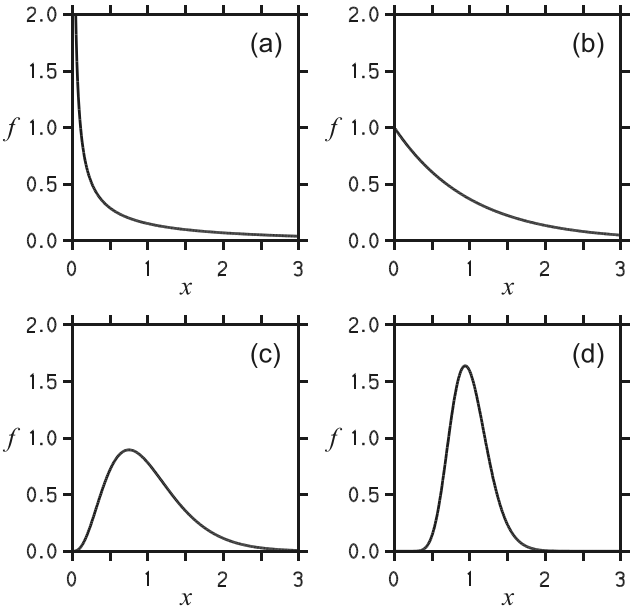



**Fig. 4.9.** An illustration of gamma PDFs for $\langle X \rangle = 1$. The standard deviation $\langle \widetilde{X}^2 \rangle^{1/2} = 2, 1, 0.5, 0.25$ in (a), (b), (c), and (d), respectively.

The obvious requirement of Eq. (4.97) is that $a = b$. Hence, we find

$$a = b = \frac{\langle X \rangle^2}{\langle \widetilde{X}^2 \rangle}, \tag{4.100}$$

which enables the calculation of $a = b$ for a given mean and variance. By using $a = b$ we can write the gamma PDF (4.88) for $x \geq 0$ as

$$f(x) = \frac{a}{\Gamma(a)\langle X \rangle} \left( \frac{a\,x}{\langle X \rangle} \right)^{a-1} \exp\left( -\frac{a\,x}{\langle X \rangle} \right). \tag{4.101}$$

**Non-Normality.** The skewness $m_3$ and flatness $m_4$ implied by the gamma PDF are given by the expressions

$$m_3 = \frac{2}{\sqrt{a}}, \qquad m_4 = 3\left(1 + \frac{2}{a}\right), \tag{4.102}$$

see exercise 4.3.9. The parameter $a$ is positive due to Eq. (4.100). Thus, for finite values of the parameter $a$ we find that both $m_3$ and $m_4$ are unequal and bigger than the corresponding values of a normal PDF. The values $m_3 = 0$ and $m_4 = 3$ for a normal PDF are recovered in the limit that $a \to \infty$.

**Illustration.** Figure 4.9 illustrates the broad range of variations of the gamma PDF. The value $a = 1$ specifies a particular PDF, the exponential PDF

$$f(x) = \frac{1}{\langle X \rangle} \exp\left( -\frac{x}{\langle X \rangle} \right), \tag{4.103}$$

for which we have $\Gamma(1) = 1$. The behavior of the gamma PDF is different for the cases that $a < 1$ and $a > 1$. For $a < 1$, the gamma PDF looks like a modified exponential function. For $a > 1$, the PDF looks like a modification of a normal PDF. The normal PDF features $m_3 = 0$ and $m_4 = 3$ are recovered for a finite mean value if the variance $< \widetilde{X}^2 >$ disappears, $< \widetilde{X}^2 > \to 0$, which implies that $a \to \infty$.

**Probability Calculation.** Unfortunately, the calculation of the probability of events cannot be performed analytically on the basis of the gamma function. A way to handle this calculation is to follow the approach used for the calculation of the integral over the normal PDF. The application of the gamma PDF (4.101) in the definition (4.16) of $P(c \le X \le d)$ leads to

$$\begin{aligned} P(c \le X \le d) &= \frac{a}{\Gamma(a)\langle X \rangle} \int_c^d \left( \frac{a\,x}{\langle X \rangle} \right)^{a-1} \exp\left( -\frac{a\,x}{\langle X \rangle} \right) dx = \frac{1}{\Gamma(a)} \int_C^D y^{a-1} e^{-y}\, dy \\ &= \frac{1}{\Gamma(a)} \left[ \int_0^D y^{a-1} e^{-y}\, dy - \int_0^C y^{a-1} e^{-y}\, dy \right] = \frac{\Gamma_I(a,D) - \Gamma_I(a,C)}{\Gamma(a)}. \end{aligned} \tag{4.104}$$

Here, we applied the substitution $y = a\,x / \langle X \rangle$ and the bounds $C = a\,c / \langle X \rangle$ and $D = a\,d / \langle X \rangle$. We did also use the definition of the incomplete gamma function

$$\Gamma_I(a,x) = \int_0^x y^{a-1} e^{-y} dy. \tag{4.105}$$

This function recovers the gamma function $\Gamma(a)$ for $x \to \infty$. The advantage of Eq. (4.104) is that we can use now the series expansion of the incomplete gamma function (Abramowitz & Stegun 1984) for the calculation of $P(c \le X \le d)$,

$$\Gamma_I(a,x) = x^a \sum_{n=0}^{\infty} \frac{(-x)^n}{(a+n)n!}. \tag{4.106}$$

### *4.3.4 The Beta Probability Density Function*

**Beta PDF.** A normal PDF can also not be applied to cases where the random variable is non-negative and has a finite range of variations. Such a case is given, e.g., regarding the mixture fraction (which represents a measure for the degree of the mixedness of a certain substance) that can only vary between zero and one by

definition. A gamma PDF cannot be used for this case, too, because it predicts nonzero probabilities for finding values of such a variable outside of the allowed range of variations. The distribution of such random variables is often described by a beta PDF that is defined by

$$f(x) = \begin{cases} \dfrac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}. \tag{4.107}$$

Here, $B(a, b)$ is the so-called beta function. This function is defined as the integral

$$B(a,b) = \int_0^1 y^{a-1}(1-y)^{b-1}\,dy. \tag{4.108}$$

$B(a, b)$ can be calculated via its relation to the gamma function (Abramowitz & Stegun 1984),

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \tag{4.109}$$

The use of the beta function definition (4.108) shows that the beta PDF integrates to one,

$$\int_0^1 f(x)\,dx = \frac{1}{B(a,b)} \int_0^1 x^{a-1}(1-x)^{b-1}\,dx = 1. \tag{4.110}$$

The characteristic features of this PDF (the moments, relations between the model parameters $a$ and $b$ with the mean and variance, the skewness and flatness) can be analyzed in correspondence to the analysis of the gamma PDF: see the exercises 4.3.10 and 4.3.11.

**Illustration.** Figure 4.10 illustrates the range of variations of the beta PDF. The beta PDF is often used to characterize the mixing of substances, so let us describe the beta PDF features with regard to this case. For relatively small values of $a = b = 0.28$, Fig. 4.10a basically describes the existence of two states: one state where the random variable $X \approx 0$, and a second state where the random variable $X \approx 1$. After a certain amount of time, mixing leads to the uniform PDF that can be seen in Fig. 4.10b. This uniform PDF is characterized by larger $a = b = 1$ values that imply $m_3 = 0$ and $m_4 = 9/5$. Additional mixing leads to the appearance of a PDF that is close to a normal PDF: see Fig. 4.10c that is obtained for $a = b = 5.1$. The corresponding flatness value $m_4 = 2.5$ is smaller than the value $m_4 = 3$ found for a normal PDF. For the case $a = b \to \infty$ we find a peak value that goes to infinity. This case corresponds to a state of perfect mixing where there is only a nonzero probability for the value $X = 0.5$. The PDF is then characterized by the values $m_3 = 0$ and $m_4 = 3$ of a normal PDF.

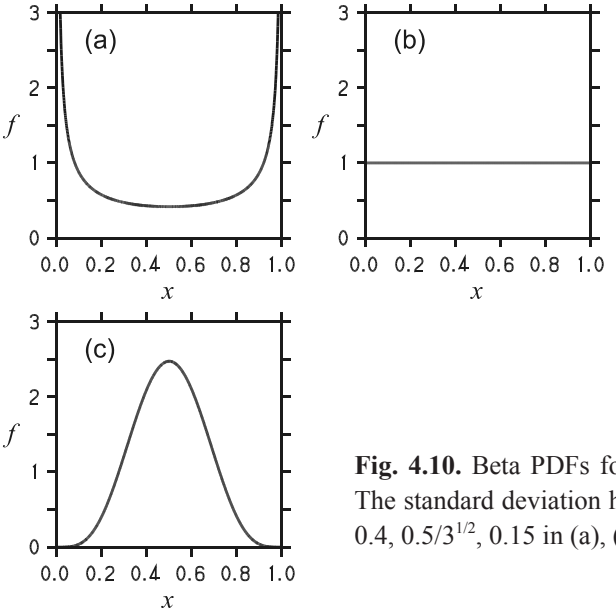



**Fig. 4.10.** Beta PDFs for a mean of $\langle X \rangle = 0.5$. The standard deviation has the values $\langle \widetilde{X}^2 \rangle^{1/2} = 0.4, 0.5/3^{1/2}, 0.15$ in (a), (b), and (c), respectively.

**Probability Calculation.** To calculate probabilities $P(c \le X \le d)$ we apply the beta PDF (4.107) in the probability definition (4.16),

$$P(c \le X \le d) = \frac{1}{B(a,b)} \int_c^d x^{a-1} (1-x)^{b-1}\, dx = \frac{1}{B(a,b)} \left[ \int_0^d x^{a-1} (1-x)^{b-1}\, dx - \int_0^c x^{a-1} (1-x)^{b-1}\, dx \right] = \frac{B_I(a,b,d) - B_I(a,b,c)}{B(a,b)}. \tag{4.111}$$

It is assumed here that $0 \le c, d \le 1$. In this equation we applied the definition of the incomplete beta function

$$B_I(a,b,x) = \int_0^x y^{a-1} (1-y)^{b-1} dy, \tag{4.112}$$

which recovers the beta function $B(a, b)$ for $x = 1$. The advantage of Eq. (4.111) is that we can apply the following series to calculate the incomplete beta function (Abramowitz & Stegun 1984),

$$B_I(a,b,x) = x^a \sum_{n=0}^{\infty} \frac{\Gamma(1-b+n)}{\Gamma(1-b)} \frac{x^n}{(a+n)n!} = x^a \sum_{n=0}^{\infty} \frac{p_n x^n}{a+n}. \tag{4.113}$$

The last expression introduces the abbreviation $p_n = \Gamma(1-b+n) / [\Gamma(1-b)\, n!]$. The definition of $p_n$ shows that $p_0 = 1$ and $p_n = p_{n-1}\, (n-b) / n$ for $n \ge 1$. Hence, $p_n$ is finite for increasing $n$. The calculation of $p_n$ by $p_n = p_{n-1}\, (n-b) / n$ enables a stable

numerical calculation of $B_I(a, b, x)$: the problem to evaluate $\Gamma(1 - b + n)$ at large $n$ (which appears if $B_I(a, b, x)$ is calculated for $x > 0.9$) is avoided in this way.

## 4.4 Data Analysis

Let us address now the question of how simulation results and measurements of random variables can be analyzed. A particular problem of such data analysis is the question of how means and PDFs derived from data depend on a finite number of samples applied.

### *4.4.1 Calculation of Statistics*

**Filtered PDFs.** The definition $f(x) = < d\theta\,(x - X) / dx >$ of a PDF involves a derivative. In order to calculate a PDF from measurements or simulation results, we have to represent the derivative in a discrete way. This can be done by using for the PDF the expression

$$\begin{aligned} f_\Delta(x) &= \frac{1}{\Delta x}\left\langle \theta\left(x + \frac{\Delta x}{2} - X\right) - \theta\left(x - \frac{\Delta x}{2} - X\right)\right\rangle \\ &= \frac{1}{\Delta x}\int_{x-\Delta x/2}^{x+\Delta x/2}\left\langle \frac{d\theta(y - X)}{dy}\right\rangle dy = \frac{1}{\Delta x}\int_{x-\Delta x/2}^{x+\Delta x/2} f(y)\,dy. \end{aligned} \tag{4.114}$$

The last expression shows that $f_\Delta(x)$ is a PDF that is filtered over the interval $\Delta x$. No assumption is made here about the interval $\Delta x$, which may be small or not. The subscript $\Delta$ refers to the filter operation defined by Eq. (4.114). This notation will be also used below for other functions that are filtered over an interval $\Delta x$. From a computational point of view it is helpful to represent $f_\Delta(x)$ in another way,

$$f_\Delta(x) = \frac{1}{\Delta x}\left\langle \theta\left(x + \frac{\Delta x}{2} - X\right) - \theta\left(x - \frac{\Delta x}{2} - X\right)\right\rangle = \frac{1}{\Delta x}\frac{\Delta N}{N}. \tag{4.115}$$

According to the definition (4.1) of means, the number $\Delta N$ is defined by

$$\begin{aligned} \Delta N &= \sum_{i=1}^{N}\left[\theta\left(x + \frac{\Delta x}{2} - X_i\right) - \theta\left(x - \frac{\Delta x}{2} - X_i\right)\right] \\ &= \sum_{i=1}^{N}\begin{Bmatrix} 1 - 1 = 0 & \text{if } X_i < x - \Delta x/2 \\ 1 - 0 = 1 & \text{if } x - \Delta x/2 \le X_i \le x + \Delta x/2 \\ 0 - 0 = 0 & \text{if } x + \Delta x/2 < X_i \end{Bmatrix}. \end{aligned} \tag{4.116}$$

Consequently, $\Delta N$ measures the number of samples that are found in the interval $x - \Delta x/2 \le X_i \le x + \Delta x/2$. The PDF $f_\Delta(x) = \Delta N/(N\,\Delta x)$ represents, therefore, the relative number of samples around $x$ normalized by the filter interval $\Delta x$.

**Properties of Filtered PDFs.** For any function $g(x)$, the filtered PDF $f_\Delta(x)$ has the property

$$\begin{aligned}\int g(x)\,f_\Delta(x)\,dx &= \frac{1}{\Delta x}\left\langle \int g(x)\left[\theta\left(x-\left[X-\frac{\Delta x}{2}\right]\right)-\theta\left(x-\left[X+\frac{\Delta x}{2}\right]\right)\right]dx\right\rangle \\ &= \frac{1}{\Delta x}\left\langle \int_{X-\Delta x/2}^{\infty} g(x)\,dx - \int_{X+\Delta x/2}^{\infty} g(x)\,dx\right\rangle = \frac{1}{\Delta x}\left\langle \int_{X-\Delta x/2}^{\infty} g(x)\,dx + \int_{\infty}^{X+\Delta x/2} g(x)\,dx\right\rangle \\ &= \frac{1}{\Delta x}\left\langle \int_{X-\Delta x/2}^{X+\Delta x/2} g(x)\,dx\right\rangle = \langle g_\Delta(X)\rangle.\end{aligned} \tag{4.117}$$

The first rewriting results from using the definition (4.114) of $f_\Delta(x)$. The following rewritings make use of the properties of theta functions. In correspondence to the expression (4.114), the last expression represents the mean of the filtered function $g_\Delta(X)$. The relation (4.117) enables some interesting observations that follow from the calculation of the integral over $g(x)$: see the first expression of the last line.

- By setting $g = 1$ we find that $f_\Delta(x)$ represents indeed a PDF because it integrates to one,

$$\int f_\Delta(x)\,dx = \frac{1}{\Delta x}\left\langle X+\frac{\Delta x}{2}-\left(X-\frac{\Delta x}{2}\right)\right\rangle = 1. \tag{4.118}$$

- By setting $g = x$ we find

$$\int x\,f_\Delta(x)\,dx = \frac{1}{2\Delta x}\left\langle \left(X+\frac{\Delta x}{2}\right)^2-\left(X-\frac{\Delta x}{2}\right)^2\right\rangle = \frac{1}{2\Delta x}\langle 2\,X\,\Delta x\rangle = \langle X\rangle. \tag{4.119}$$

- By setting $g = (x - \langle X\rangle)^2$ we obtain

$$\begin{aligned}\int \left(x-\langle X\rangle\right)^2 f_\Delta(x)\,dx &= \frac{1}{3\Delta x}\left\langle \left(X-\langle X\rangle+\frac{\Delta x}{2}\right)^3-\left(X-\langle X\rangle-\frac{\Delta x}{2}\right)^3\right\rangle \\ &= \frac{1}{3\Delta x}\Big\langle \widetilde{X}^3+3\,\widetilde{X}^2\,\Delta x/2+3\,\widetilde{X}\,(\Delta x/2)^2+(\Delta x/2)^3 \\ &\qquad -\widetilde{X}^3+3\,\widetilde{X}^2\,\Delta x/2-3\,\widetilde{X}\,(\Delta x/2)^2+(\Delta x/2)^3\Big\rangle \\ &= \frac{\left\langle 6\,\widetilde{X}^2\,\Delta x/2+2\,(\Delta x/2)^3\right\rangle}{3\Delta x} = \langle \widetilde{X}^2\rangle+\frac{1}{12}(\Delta x)^2.\end{aligned} \tag{4.120}$$

What is the relevance of these properties of filtered PDFs? The Eqs. (4.118) and (4.119) show that the filtered PDF integrates correctly to one, and it provides the correct mean value. Equation (4.120) shows that the variance of the filtered PDF

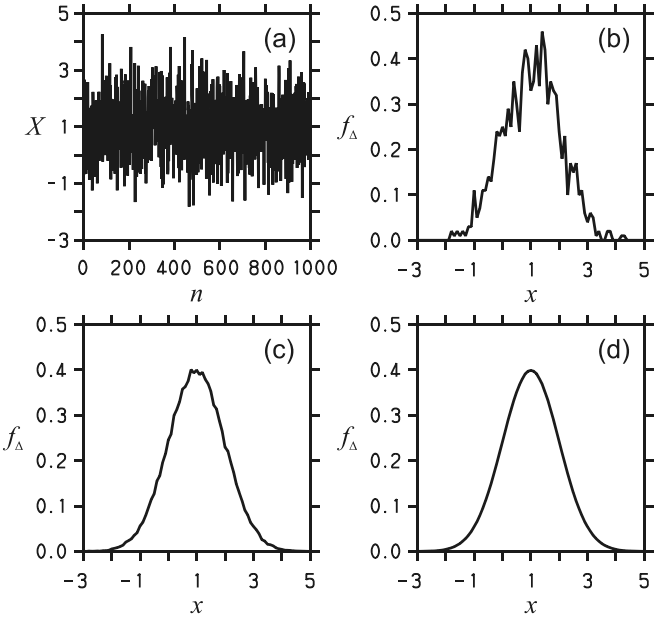



**Fig. 4.11.** The sample number effect on the PDF $f_\Delta(x)$. (a) The data set considered is illustrated for $N = 10^3$; (b), (c), and (d) show $f_\Delta(x)$ for $N = 10^3$, $N = 10^5$, $N = 10^7$, respectively, where $\Delta x = 0.1$ is used.

is always bigger than the real variance of the data set. The deviation $(\Delta x)^2 / 12$ can be used to assess the accuracy of the filtered PDF. Consequently, the use of the filtered PDF for the calculation of probabilities can only be approximately correct depending on the relevance of $(\Delta x)^2 / 12$. For the case that $(\Delta x)^2 / 12$ is not small, it is possible to use the definition

$$P(a \le X \le b) = \langle \theta(b - X) - \theta(a - X) \rangle = \frac{\Delta N_{ab}}{N} \tag{4.121}$$

for the calculation of correct probabilities. The expression $\Delta N_{ab} / N$ follows from Eq. (4.115) by setting $a = x - \Delta x / 2$ and $b = x + \Delta x / 2$: this means $\Delta N_{ab}$ refers to the number of samples in the interval between $a$ and $b$.

**Sample Number Effect.** The calculation of PDFs in dependence on the sample number $N$ is illustrated in Fig. 4.11. The data set applied is illustrated for $N = 10^3$ in Fig. 4.11a. The data represent normally distributed random numbers with zero mean and a standard deviation of one. The PDFs in Figs. 4.11b–d are calculated in terms of the definition $f_\Delta(x) = \Delta N / (N \Delta x)$ of a filtered PDF for a fixed interval $\Delta x = 0.1$ for varying sample numbers $N$. The range of $x$ values considered covers the range of variations of $X$ values – see Fig. 4.11a. The positions $x$ at which the

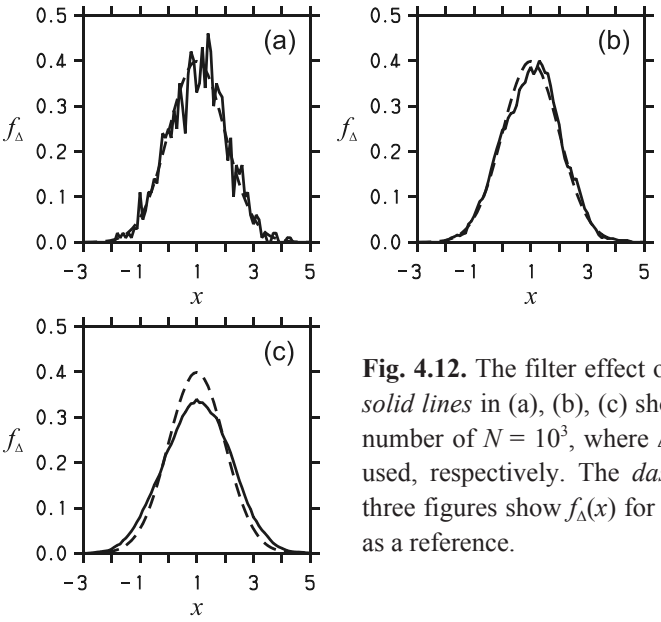



**Fig. 4.12.** The filter effect on the PDF $f_\Delta(x)$. The *solid lines* in (a), (b), (c) show $f_\Delta(x)$ for a sample number of $N = 10^3$, where $\Delta x = (0.1, 0.5, 2)$ are used, respectively. The *dashed lines* in all the three figures show $f_\Delta(x)$ for $N = 10^7$ and $\Delta x = 0.1$ as a reference.

PDF is calculated are separated by 0.1. Figure 4.11b reveals the problem of calculating a PDF on the basis of a relatively low number $N = 10^3$ of samples: the PDF is significantly affected by noise. An increasing number of samples results in a much smoother PDF. The PDF for $N = 10^7$ samples agrees exactly with the corresponding normal PDF, which means that a further increase of the sample number does not cause any modifications of the PDF.

**Filter Interval Effect.** For a relatively low number $N$ of samples, the need to work with relatively smooth PDFs requires the use of a relatively large filter width $\Delta x$ to have a sufficient number of samples in the intervals. Figure 4.12 illustrates the effect of $\Delta$x variations for PDFs with a number $N = 10^3$ of samples. The PDF with $N = 10^7$ and $\Delta x = 0.1$ is given as a reference. These figures illustrate again the problem of calculating PDFs from a data set with a limited number of samples. For a small interval $\Delta x = 0.1$ the calculated PDF is significantly influenced by noise due to the low number of samples in the intervals. For a large interval $\Delta x = 2$ the PDF is a relatively smooth function. However, the large filter width $\Delta x$ applied leads to the fact that the PDF is smeared out: the peak value is signifi-cantly underpredicted and the probability for high deviations from the mean value is overpredicted. There is no unique solution for determining an optimal filter width: $\Delta x$ has to be large enough such that the noise effect is damped out but sufficiently small such that the PDF is not smeared out too much. For the case considered, the filtered PDF with $\Delta x = 0.5$ represents such a reasonable choice.

### *4.4.2 The First Fundamental Theorem of Probability*

**The Problem with Randomness.** The problem related to the analysis of random numbers, which can be obtained from measurements or models that involve randomness, is that every calculation of statistics does provide different results in general. For example, the calculation of a mean value

$$\mu_N = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{4.122}$$

based on a finite number $N$ of samples will provide different results depending on the number $N$ of sample values applied (and the set of sample values $X_1, X_2, \ldots X_N$ used for a given $N$). It is often the case that there is no way to change this situation because there are no other data. However, what we would like to know is under which conditions we will have exact results that can be reproduced. This is relevant to the understanding of the standard of results: it makes a difference to know that a particular result is just random, or to know that our result is very close to the exact result, or to know that we have an exact result.

**The Law of Large Numbers.** After 20 years of analyzing the latter problem, Jacob Bernoulli published in 1713 an answer to this relevant question. He named his conclusion the "Golden Theorem". In 1835, Siméon Denis Poisson introduced the name "The Law of Large Numbers" for Bernoulli's result. Bernoulli's theorem states that an infinite sequence of independent and identically distributed random numbers $X_1, X_2, \ldots$ (the random numbers are independent and each variable has the same PDF) with finite mean converges to a mean $<X>$ (Ross 2010),

$$\lim_{N\to\infty} \mu_N = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} X_i = \langle X \rangle. \tag{4.123}$$

The assumption of a finite variance is not necessary. The relevance of Bernoulli's theorem is the conclusion that the analysis of random numbers will result in exact conclusions that can be reproduced provided the number of samples is huge.

**Illustration.** In place of a proof of the Law of Large Numbers, which may be found, e.g., in Ross (2010), we consider a numerical illustration of the correctness of this law. In particular, we extend this illustration by considering not only the mean but also the standard deviation of normally distributed random numbers with $<X> = 1$ and standard deviation $< \widetilde{X}^2 >^{1/2} = 1$ for a varying number $N$ of samples. The deviations from the exact values $<X> = 1$ and $< \widetilde{X}^2 >^{1/2} = 1$ will be assessed in terms of the relative errors

$$\Delta_m = \frac{\mu_N - \langle X \rangle}{\langle X \rangle}, \qquad \Delta_{sd} = \frac{\sigma_N - \langle \widetilde{X}^2 \rangle^{1/2}}{\langle \widetilde{X}^2 \rangle^{1/2}}. \tag{4.124}$$

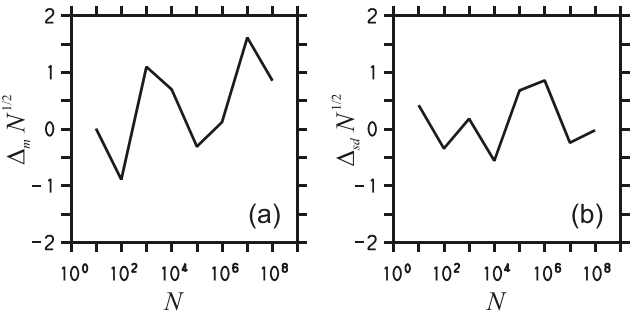



**Fig. 4.13.** (a) The relative error $\Delta_m N^{1/2}$ of the mean value multiplied with the square root of the number $N$ of samples. The corresponding relative error $\Delta_{sd} N^{1/2}$ of the standard deviation is shown in (b).

Here, $\sigma_N = <\widetilde{X}^2>_N^{1/2}$, where the fluctuation $\widetilde{X} = X - \mu_N$ refers to deviations from the mean $\mu_N$ obtained for a finite number of samples. The relative errors (4.124) multiplied with $N^{1/2}$ are shown in Fig. 4.13 in dependence on $N$. This figure shows that $\Delta_m$ and $\Delta_{sd}$ can be described by the relation

$$\Delta_m = \Delta_{sd} = \frac{\varepsilon}{N^{1/2}}, \tag{4.125}$$

where $\varepsilon$ is a bounded random variable (Fig. 4.13 suggests that $|\varepsilon| < 2$). Hence, $\mu_N$ and the standard deviation $\sigma_N$ converge to their exact values $<X>$ and $<\widetilde{X}^2>^{1/2}$ proportional to $N^{-1/2}$. Equation (4.125) can be used to specify the error of using a finite number $N$ of samples. For example, we find $|\Delta_m| = |\Delta_{sd}| < 0.02$ and $|\Delta_m| = |\Delta_{sd}| < 0.002$ for $N = 10^4$ and $N = 10^6$, respectively.

**Unbiased Estimates.** A closely related question is the following one. Usually, there is only a finite number of samples that have to be used to find estimates for means and variances. Such estimates are still random numbers. A desired property of such estimates is that these estimates represent unbiased estimates, this means estimates that have mean values that agree with the exact means and variances given for an infinite number of samples. The sample mean $\mu_N$ (4.122) represents such an unbiased estimate because

$$\langle \mu_N \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle X_i \rangle = \frac{N \langle X \rangle}{N} = \langle X \rangle. \tag{4.126}$$

Here, we replaced $<X_i>$ by $<X>$ because $X_i$ is just any random realization. An unbiased estimate for the variance is given by

$$\sigma^2{}_N = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \mu_N)^2. \tag{4.127}$$

This formula differs from the sample mean formula (4.122) by the replacement of $N$ by $N - 1$, which is called the Bessel correction. Bessel's correction is correct because there are only $N - 1$ independent samples available, $(X_1 - \mu_N, X_2 - \mu_N, \cdots, X_N - \mu_N)$, as these samples sum to zero. It is interesting that Eq. (4.127) corrects the bias in the estimation of variances, but the corresponding standard deviation $\sigma_N$ is biased. A general formula for an unbiased standard deviation does not exist.

### *4.4.3 The Second Fundamental Theorem of Probability*

**The Generalized Problem with Randomness.** The Law of Large numbers addresses the question of what is the limiting mean value of a sum $N\,\mu_N = X_1 + \cdots + X_N$ of $N$ independent and identically distributed random numbers as $N \to \infty$. A more general question is the following one: what is the limiting behavior of the PDF of a sum $N\,\mu_N = X_1 + \cdots + X_N$ of $N$ independent and identically distributed random numbers as $N$ approaches infinity? An answer to this question is certainly helpful. A sum of independent and identically distributed contributions represents a reasonable model for random numbers. Thus, the PDF of such random numbers may be expected to be a reasonable model for random variables.

**The Central Limit Theorem.** A first answer to this question was presented in 1773 by Abraham de Moivre. In 1812, Pierre-Simon Laplace expanded Moivre's findings. The result of Laplace's analysis can be summarized in the following way: Let $X_1, X_2, \cdots, X_N$ be a sequence of independent and identically distributed random numbers each having a finite mean $\mu$ and a finite variance $\sigma^2 > 0$. Further, let $D_N$ be the PDF of the sum of $N$ values $X_i$. Then, the Central Limit Theorem says that $D_N$ converges independent of the original PDF to a normal distribution with mean $N\,\mu$ and variance $N\,\sigma^2$,

$$\sum_{i=1}^{N} X_i \sim D_N \quad \Rightarrow \quad \lim_{N\to\infty} D_N = \mathcal{N}\left(N\,\mu, N\,\sigma^2\right). \tag{4.128}$$

The symbol $\mathcal{N}(M, V)$ used here refers to a normal distribution with mean $M$ and variance $V$. Alternatively, the Central Limit Theorem can be also presented for the PDF $d_N$ of mean values $\langle X \rangle_N$,

$$\frac{1}{N}\sum_{i=1}^{N} X_i \sim d_N \quad \Rightarrow \quad \lim_{N\to\infty} d_N = \mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right). \tag{4.129}$$

These two statements of the Central Limit Theorem are in consistency with the conclusions (10.75) and (10.76) for the sum and the mean of normally distributed random numbers, respectively.

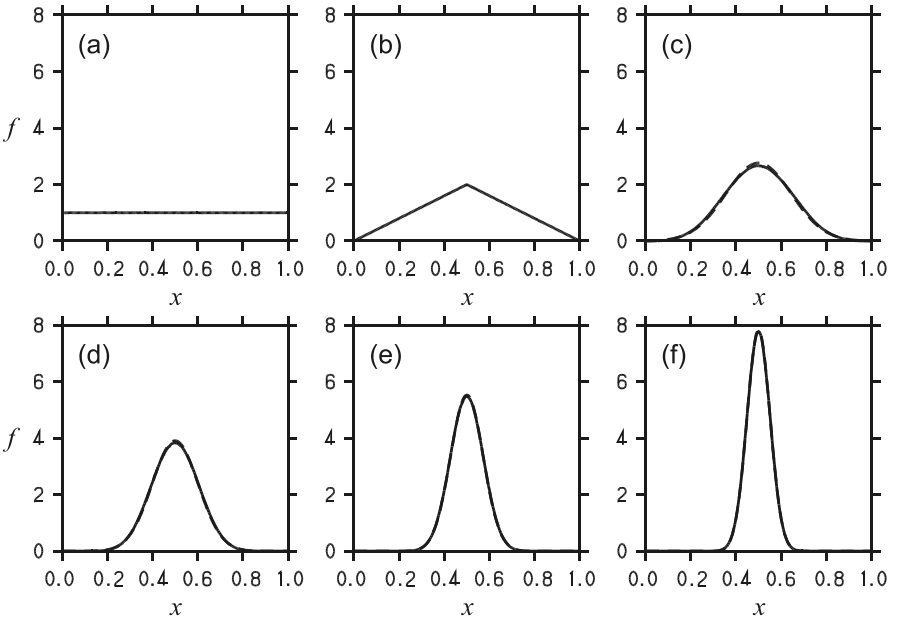



**Fig. 4.14.** The Central Limit Theorem. (a) The initial uniform PDF. The other figures show PDFs of the means of 2, 4, 8, 16, 32 samples (*solid lines*). Corresponding normal PDFs are shown in (c), (d), (e), and (f) by *dashed lines*.

**Illustration.** The Central Limit Theorem is illustrated in Figs. 4.14 and 4.15. The initial PDFs in these figures show the PDFs of $N = 10^8$ random numbers. The PDFs are calculated according to $f_\Delta(x) = \Delta N / [N \Delta x]$ with $\Delta x = 0.005$. The initial PDF in Fig. 4.14 represents a uniform PDF given by Eq. (4.33), where $a = 0$ and $b = 1$. The initial PDF in Fig. 4.15 represents a beta PDF given by (4.107), where $a = b = 0.3$. The following PDFs in these two figures represent the PDFs of mean values of 2, 4, 8, 16, 32 random numbers, respectively. The sequences of PDFs in these plots describe an increased mixing between the values of random numbers. The initial PDFs represent two sorts of unmixed random numbers: the uniform PDF defines an equal probability for all states between zero and one, and the beta PDF considers only two basic states (one state where the random variable $X \approx 0$, and another state where the random variable $X \approx 1$). The mean values of 2, 4, 8, 16, 32 random numbers characterize an increased mixing intensity of random numbers. Therefore, the corresponding PDFs are characterized by a growing peak value and a decreasing variance. These PDF shapes indicate decreasing deviations of the mean value $\langle X \rangle = 0.5$ with a growing number of samples involved in the mean values. In agreement with the Central Limit Theorem, we find independent of the initial PDF decreasing deviations between these PDFs and a normal PDF

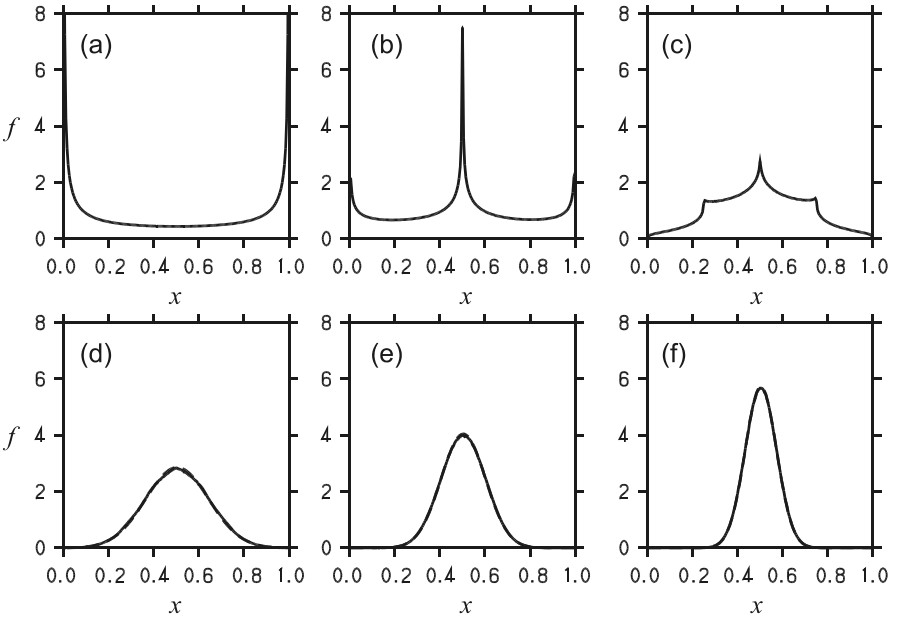



**Fig. 4.15.** The Central Limit Theorem. (a) The initial beta PDF. The other figures show PDFs of the means of 2, 4, 8, 16, 32 samples (*solid lines*). Corresponding normal PDFs are shown in (d), (e), and (f) by *dashed lines*.

for an increasing number $N$ of sample values involved. According to the Central Limit Theorem (4.129), the normal PDFs are given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_0^{\,2}/N}} \exp\left\{-\frac{(x-\mu_0)^2}{2\sigma_0^{\,2}/N}\right\}. \tag{4.130}$$

Here, $\mu_0$ and $\sigma_0^2$ are the mean and variance of the initial PDFs. For the uniform initial PDF, we have the values $\mu_0 = 0.5$ and $\sigma_0^2 = 1/12$. Regarding the beta initial PDF, we have $\mu_0 = 0.5$ and $\sigma_0^2 = 13/32$.

**Entropy.** The PDF for the mean of eight samples is already very close to a normal PDF. The difference to the normal PDF can be quantitatively assessed by calculating the entropy according to Eq. (4.53). As shown above, a normal PDF has a higher entropy than any other PDF with the same mean and variance. The latter fact is illustrated in Fig. 4.16 that shows the entropy $S$ for the PDFs given in Figs. 4.14 and 4.15 in comparison to the entropy $S = \ln(2\,e\,\pi\,\sigma^2)^{1/2}$ of normal PDFs. Also, the entropy of non-normal PDFs is always smaller than the entropy of the normal PDFs. These figures also reveal that the non-normal PDFs quickly converge to the corresponding normal PDFs. For PDFs for the mean of eight samples there is no observable difference between the entropy values.

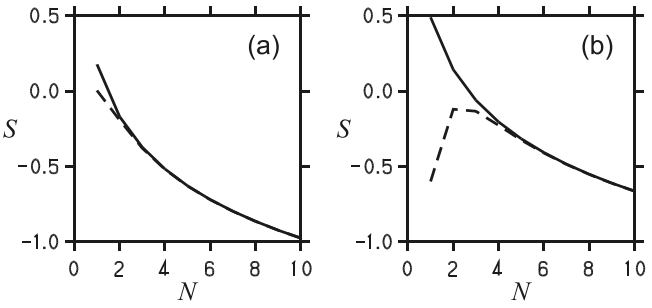



**Fig. 4.16.** The entropy is shown by a *dashed line* in (a) with regard to the Fig. 4.14 PDFs and in (b) with regard to the Fig. 4.15 PDFs. The entropy of the corresponding normal PDFs is given by *solid lines*.

## 4.5 Real Distributions

Let us consider now the PDFs derived from real data and the suitability of PDF modeling concepts discussed above. In particular, we will consider the data of velocity components and temperature that were measured by Chu et al. (1996) in the atmospheric surface layer for different stability conditions. Such PDFs are of interest to our daily life, for example, for the assessment of the probability for finding any (extreme) events. Also, such PDFs are used for the evaluation of atmospheric turbulence models, and they provide a basic ingredient for the development of models for the transport of species in the atmosphere (Luhar et al. 1996).

### *4.5.1 Velocity and Temperature Data*

**Atmospheric Stability.** Measurements for different atmospheric stabilities (in particular for stable, neutral, and unstable conditions) will be considered below. Let us briefly describe these conditions to prepare these discussions. Atmospheric stability is the degree to which the atmosphere will support, tolerate, or suppress vertical motions. In a stable atmosphere, a parcel of air that is displaced upwards will tend to return to its original level. In an unstable atmosphere, a parcel of air displaced upwards will continue to rise. Stable conditions typically occur during nighttime or in winter situations. Unstable conditions typically occur during daytime at summer when the Sun is shining. Neutral conditions typically occur in the transition from day to night.

**Table 4.1** Statistical characteristics of the velocity and temperature PDFs.

| | | $\langle X \rangle$ | $\langle \tilde{X}^2 \rangle$ | $m_3$ | $m_4$ | $m_5$ | $m_6$ |
|---|---|---|---|---|---|---|---|
| **Stable Case:** | $u$ | 2.89 | 0.17 | 0.17 | 3.02 | 2.93 | 17.66 |
| | $v$ | 0.00 | 0.14 | –0.37 | 2.27 | –2.30 | 8.54 |
| | $w$ | 0.04 | 0.02 | 0.19 | 5.47 | 2.36 | 78.44 |
| | $T$ | 21.89 | 0.36 | 0.09 | 2.64 | 0.89 | 10.96 |
| **Neutral Case:** | $u$ | 6.94 | 0.54 | 0.08 | 3.11 | 1.05 | 17.49 |
| | $v$ | 0.00 | 0.28 | –0.05 | 3.41 | –0.43 | 21.03 |
| | $w$ | 0.22 | 0.13 | 0.14 | 3.77 | 1.83 | 29.10 |
| | $T$ | 31.47 | 0.23 | 0.56 | 3.08 | 4.65 | 17.20 |
| **Unstable Case:** | $u$ | 1.87 | 0.44 | –0.01 | 2.27 | 0.13 | 7.12 |
| | $v$ | 0.00 | 0.35 | –0.01 | 2.86 | 0.32 | 13.11 |
| | $w$ | 0.01 | 0.07 | 0.32 | 3.13 | 2.86 | 17.57 |
| | $T$ | 34.80 | 0.49 | 1.19 | 5.10 | 16.27 | 66.43 |

**Measurements.** Velocity and temperature measurements were performed at a height between 2 and 3 m above the surface. For each stability condition, 50,400 velocity and temperature values were measured over a recording period of 15 min. The short sampling period was needed to ensure steady state in the mean meteorological conditions. Each data set contains the longitudinal velocity component $u$, lateral velocity component $v$, vertical velocity component $w$, and air temperature $T$. The velocities were measured in m / s and the temperature in °C. All the details about the velocity and temperature data can be found elsewhere (Chu et al. 1996).

**PDF Calculation.** The velocity and temperature PDFs obtained from these data are shown in Fig. 4.17. The PDFs were calculated as described in Sect. 4.4.1. A relatively small filter interval $\Delta x = 0.2$ was used to ensure that the filtering does only have a little influence on the calculated PDF. The application of a larger filter interval $\Delta x$ would result, for example, in a reduction of the PDF maximum. The PDF values were calculated at sample space values that are separated by 0.02. Typical characteristics of these PDFs are given in Table 4.1. Here, $\langle X \rangle$ and $\langle \tilde{X}^2 \rangle$ refer to the mean and variance of the variables considered, respectively. The variables $m_3$, $m_4$, $m_5$, and $m_6$ represent normalized standardized central moments of third, fourth, fifth, and sixth order,

$$m_3 = \frac{\left\langle \tilde{X}^3 \right\rangle}{\left\langle \tilde{X}^2 \right\rangle^{3/2}}, \qquad m_4 = \frac{\left\langle \tilde{X}^4 \right\rangle}{\left\langle \tilde{X}^2 \right\rangle^{2}}, \qquad m_5 = \frac{\left\langle \tilde{X}^5 \right\rangle}{\left\langle \tilde{X}^2 \right\rangle^{5/2}}, \qquad m_6 = \frac{\left\langle \tilde{X}^6 \right\rangle}{\left\langle \tilde{X}^2 \right\rangle^{3}}. \quad (4.131)$$

The superskewness $m_5$ and superflatness $m_6$ values are included in Table 4.1 in addition to the skewness $m_3$ and flatness $m_4$ values because these variables will be used below in Sect. 4.5.3.

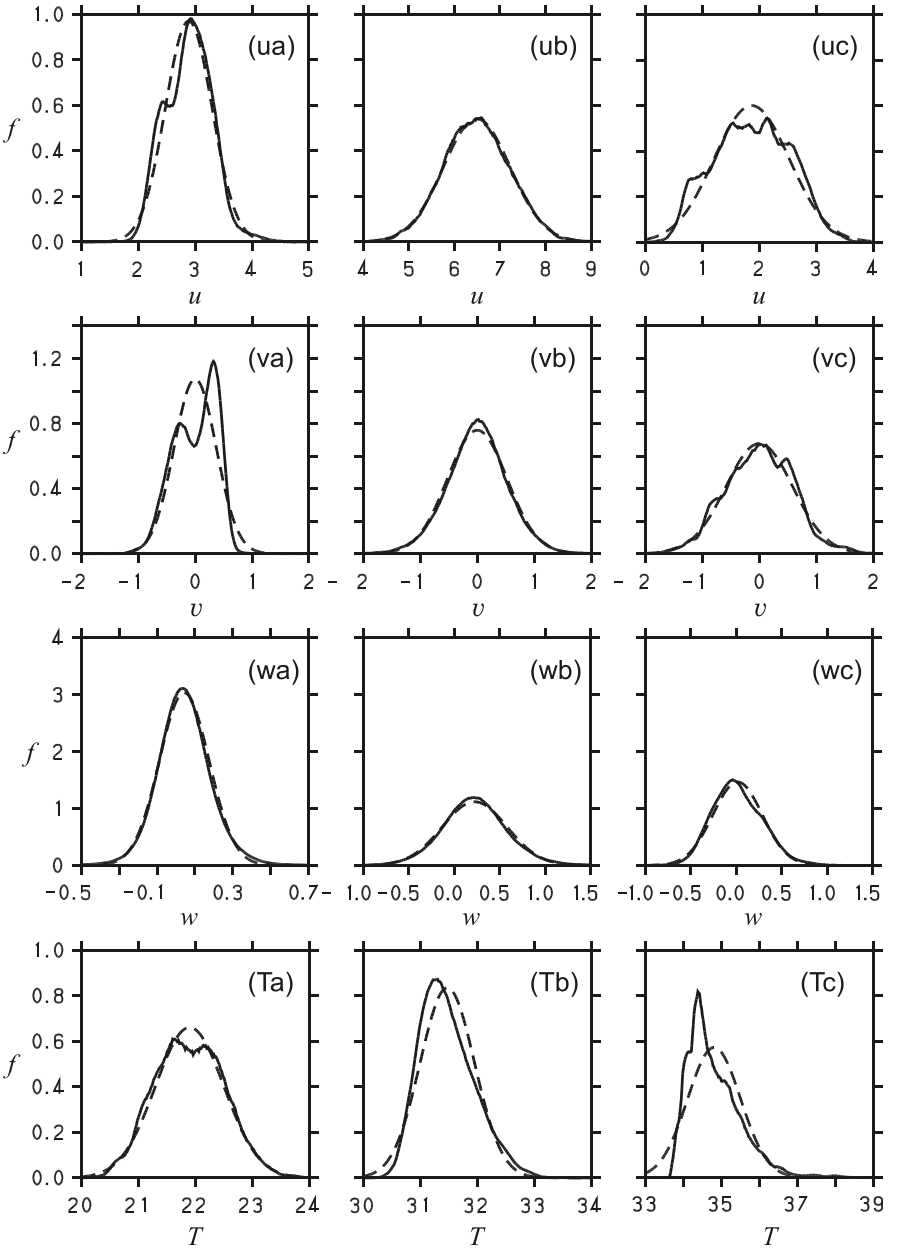



**Fig. 4.17.** The *solid lines* show velocity ($u$, $v$, $w$) and temperature ($T$) PDF data for stable (*left-hand side*), neutral (*middle*), and unstable (*right-hand side*) cases. The *dashed lines* show the corresponding normal PDFs.

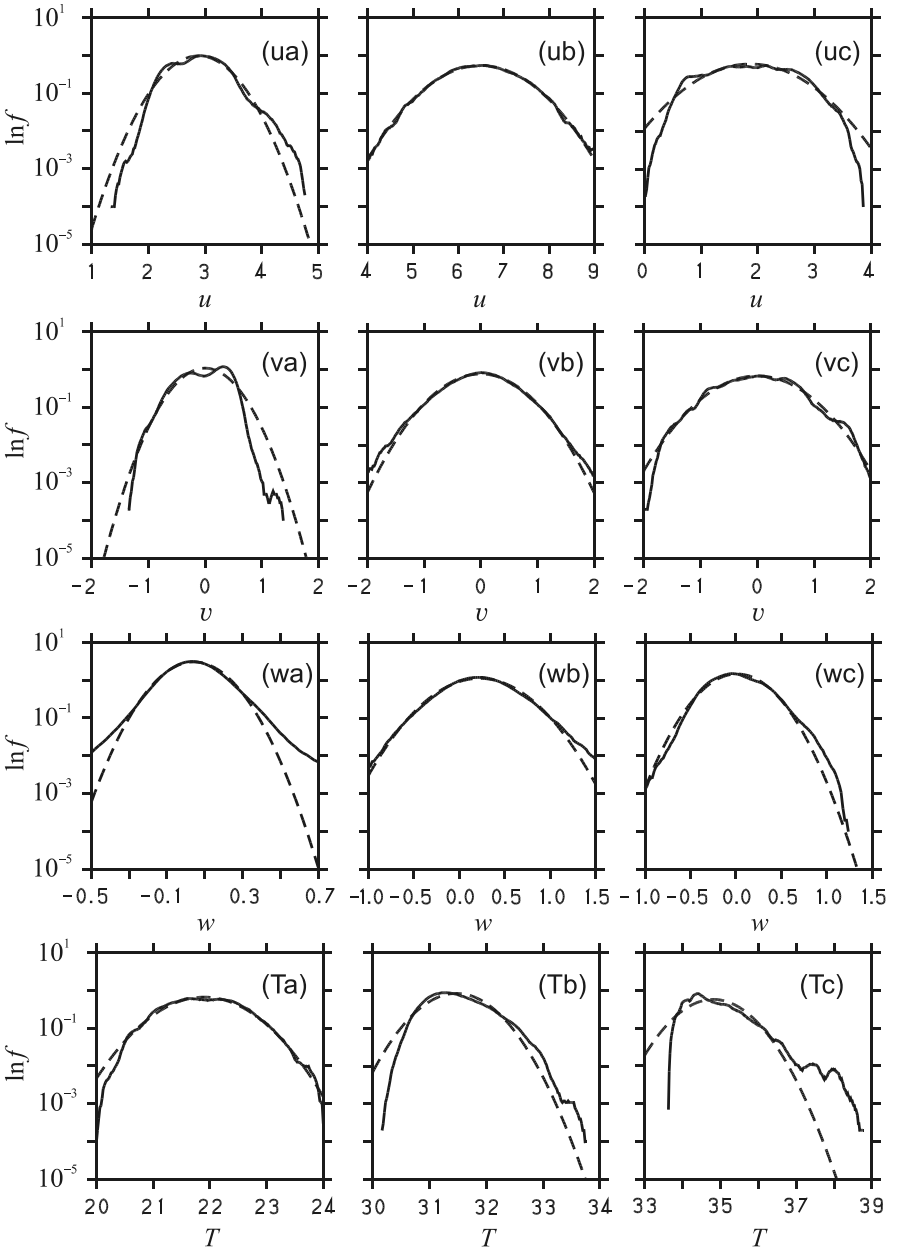



**Fig. 4.18.** The *solid lines* show the natural logarithm of velocity ($u$, $v$, $w$) and temperature ($T$) PDF data for stable (*left-hand side*), neutral (*middle*) and unstable (*right-hand side*) cases. The *dashed lines* show $\ln f$ of the normal PDFs.

### *4.5.2 Normal Probability Density Functions*

**Normal Velocity PDFs.** According to the Central Limit Theorem, the natural way to model PDFs derived from measurements is the use of normal PDFs. The mean values and variances required for the calculation of such normal PDFs are given in Table 4.1. Figure 4.17 illustrates the performance of normal PDFs. The three velocity PDFs can be described in a reasonable way by a normal PDF. The best agreement between the normal PDF and measured PDF is given for neutral conditions. This finding is not surprising because of the lack of forces that affect air parcels: the randomness under neutral conditions can be seen as purely chaotic. The biggest deviation between the normal PDF model and the measured PDFs is given regarding the $u$-PDF and $v$-PDF under stable conditions. Oscillations of air parcels about the equilibrium position under stable conditions imply that these turbulent motions are not completely random. The existence of such deterministic oscillations leads to the appearance of two different modes, see, in particular, the bimodal structure of the $v$-PDF.

**Temperature PDFs.** The normal PDF represents a reasonable model for the temperature PDF under stable conditions, but the performance of the normal PDF under neutral and unstable conditions is not very good. Under unstable conditions, the temperature PDF can be considered to represent a superposition of two modes. There is a first mode related to relatively warm particles that move up. This mode has a relatively large variance and, therefore, a relatively low peak value. There is also a second mode related to relatively cold particles that move down. This mode has a smaller variance and a higher peak value than the first mode. Mathematically, these modes imply a high skewness value $m_3 = 1.19$ under unstable conditions. A normal PDF, which has a zero skewness value, cannot represent a good model for such conditions. It is interesting to note that the temperature PDF under neutral conditions also has a significant skewness of $m_3 = 0.56$. A possible reason for this finding is that the conditions were not truly neutral but affected by instability.

**Problem Analysis.** Which model can describe the measured PDFs in a better way? A look at the non-normal PDF models described in Sect. 4.3 shows that these models do not represent alternative choices. It would be possible to use the gamma PDF for the modeling of the skewed PDFs, but such a model does not represent a general model because it cannot properly describe PDFs that are close to a normal PDF. To see how an improved model can be designed, it is helpful to consider $\ln f$ that is shown in Fig. 4.18. The PDFs can be described by a normal PDF if $\ln f$ can be approximated by a quadratic function. The comparison with the corresponding normal PDFs shows that the latter is not always possible with a reasonable accuracy. A way to improve the PDF model performance is to use a curve that offers a higher flexibility than a quadratic function, this means a poly-

nomial of an order higher than two. The application of a polynomial of third order results in a PDF model where the PDF diverges at high positive or negative values of the sample space variable (a third-order polynomial will approach infinity for large positive or negative values of the sample space variable such that the corresponding PDF diverges). Therefore, the next better approximation compared to the use of a second-order polynomial is the use of a fourth-order polynomial for the modeling of ln *f*. The development of such a PDF model will be described next.

### *4.5.3 Statistically Most-Likely Probability Density Functions*

**Fourth-Order SML PDF.** According to the discussion in the preceding paragraph we consider a fourth-order SML PDF. This PDF has a maximal entropy *S* (uncertainty) among all PDFs for which the first four moments agree with the first four moments of any data set. Equation (4.59) shows that this PDF is given by the expression

$$f(x) = \exp\left\{\lambda_0 - \lambda_1 \frac{x - \langle X\rangle}{\langle \tilde{X}^2\rangle^{1/2}} - \frac{\lambda_2}{2}\frac{(x - \langle X\rangle)^2}{\langle \tilde{X}^2\rangle} - \frac{\lambda_3}{3}\frac{(x - \langle X\rangle)^3}{\langle \tilde{X}^2\rangle^{3/2}} - \frac{\lambda_4}{4}\frac{(x - \langle X\rangle)^4}{\langle \tilde{X}^2\rangle^{2}}\right\}. \tag{4.132}$$

The use of this model requires the determination of the five model parameters $\lambda_0$, $\lambda_1$, $\lambda_2$, $\lambda_3$, and $\lambda_4$. There are five conditions for the calculation of these parameters given by the constraint that $f(x)$ integrates to one, and the four conditions that the first four moments of $f(x)$ agree with the first four moments of any data set.

**Model Parameter Calculation.** It is convenient to differentiate $f(x)$ in order to simplify the calculation of the model parameters,

$$\frac{df}{dx} = \frac{f}{\langle \tilde{X}^2\rangle^{1/2}}\left\{-\lambda_1 - \lambda_2 \frac{x - \langle X\rangle}{\langle \tilde{X}^2\rangle^{1/2}} - \lambda_3 \frac{(x - \langle X\rangle)^2}{\langle \tilde{X}^2\rangle} - \lambda_4 \frac{(x - \langle X\rangle)^3}{\langle \tilde{X}^2\rangle^{3/2}}\right\}. \tag{4.133}$$

By multiplying this relation with $< \tilde{X}^2 >^{1/2} / \lambda_2$ and writing the term that involves the product $f\,(x - <X>)$ on the left-hand side we obtain

$$\frac{x - \langle X\rangle}{\langle \tilde{X}^2\rangle^{1/2}} f = -\frac{f}{\lambda_2}\left\{\lambda_1 + \lambda_3 \frac{(x - \langle X\rangle)^2}{\langle \tilde{X}^2\rangle} + \lambda_4 \frac{(x - \langle X\rangle)^3}{\langle \tilde{X}^2\rangle^{3/2}}\right\} - \frac{\langle \tilde{X}^2\rangle^{1/2}}{\lambda_2}\frac{df}{dx}. \tag{4.134}$$

The multiplication of this relation with appropriate powers of $(x - <X>) / < \widetilde{X}^2 >^{1/2}$ and integration then provides the conditions

$$0 = \frac{1}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}} \int \left(x - \left\langle X \right\rangle\right) f \, dx = -\frac{\lambda_1 + \lambda_3 + \lambda_4 m_3}{\lambda_2}, \tag{4.135a}$$

$$1 = \frac{1}{\left\langle \widetilde{X}^2 \right\rangle} \int \left(x - \left\langle X \right\rangle\right)^2 f \, dx = -\frac{\lambda_3 m_3 + \lambda_4 m_4 - 1}{\lambda_2}, \tag{4.135b}$$

$$m_3 = \frac{1}{\left\langle \widetilde{X}^2 \right\rangle^{3/2}} \int \left(x - \left\langle X \right\rangle\right)^3 f \, dx = -\frac{\lambda_1 + \lambda_3 m_4 + \lambda_4 m_5}{\lambda_2}, \tag{4.135c}$$

$$m_4 = \frac{1}{\left\langle \widetilde{X}^2 \right\rangle^{2}} \int \left(x - \left\langle X \right\rangle\right)^4 f \, dx = -\frac{\lambda_1 m_3 + \lambda_3 m_5 + \lambda_4 m_6 - 3}{\lambda_2}, \tag{4.135d}$$

where $m_3$, $m_4$, $m_5$, and $m_6$ are defined by Eq. (4.131). Partial integration has to be applied to treat terms involving $df / dx$. The conditions (4.135) can be also written

$$\lambda_1 + \lambda_3 + \lambda_4 m_3 = 0, \tag{4.136a}$$

$$\lambda_2 + \lambda_3 m_3 + \lambda_4 m_4 = 1, \tag{4.136b}$$

$$\lambda_1 + \lambda_2 m_3 + \lambda_3 m_4 + \lambda_4 m_5 = 0, \tag{4.136c}$$

$$\lambda_1 m_3 + \lambda_2 m_4 + \lambda_3 m_5 + \lambda_4 m_6 = 3. \tag{4.136d}$$

Equations (4.136a–b) can be used to express $\lambda_1$ and $\lambda_2$ in terms of $\lambda_3$ and $\lambda_4$,

$$\lambda_1 = -\lambda_3 - \lambda_4 m_3, \qquad \lambda_2 = 1 - \lambda_3 m_3 - \lambda_4 m_4. \tag{4.137}$$

We use these relations for replacing $\lambda_1$ and $\lambda_2$ such that (4.136c–d) read

$$\begin{aligned} &-\lambda_3 - \lambda_4 m_3 + (1 - \lambda_3 m_3 - \lambda_4 m_4) m_3 + \lambda_3 m_4 + \lambda_4 m_5 = a\lambda_3 + b\lambda_4 + m_3 = 0, \\ &-(\lambda_3 + \lambda_4 m_3) m_3 + (1 - \lambda_3 m_3 - \lambda_4 m_4) m_4 + \lambda_3 m_5 + \lambda_4 m_6 = b\lambda_3 + c\lambda_4 + m_4 = 3. \end{aligned} \tag{4.138}$$

Here, the abbreviations $a$, $b$, and $c$ are given by the expressions

$$a = m_4 - m_3^{\,2} - 1, \qquad b = m_5 - m_3 m_4 - m_3, \qquad c = m_6 - m_4^{\,2} - m_3^{\,2}. \tag{4.139}$$

The solutions of the linear equation system (4.138) are then given by

$$\lambda_3 = \frac{m_3 c + (3 - m_4) b}{b^2 - ac}, \qquad \lambda_4 = -\frac{m_3 b + (3 - m_4) a}{b^2 - ac}. \tag{4.140}$$

The values of $\lambda_1$ and $\lambda_2$ can be obtained by using the latter relations for $\lambda_3$ and $\lambda_4$. The parameter $\lambda_0$ that is not determined by these relations can be calculated finally by the constraint that the model (4.132) does satisfy the normalization condition to integrate to one.

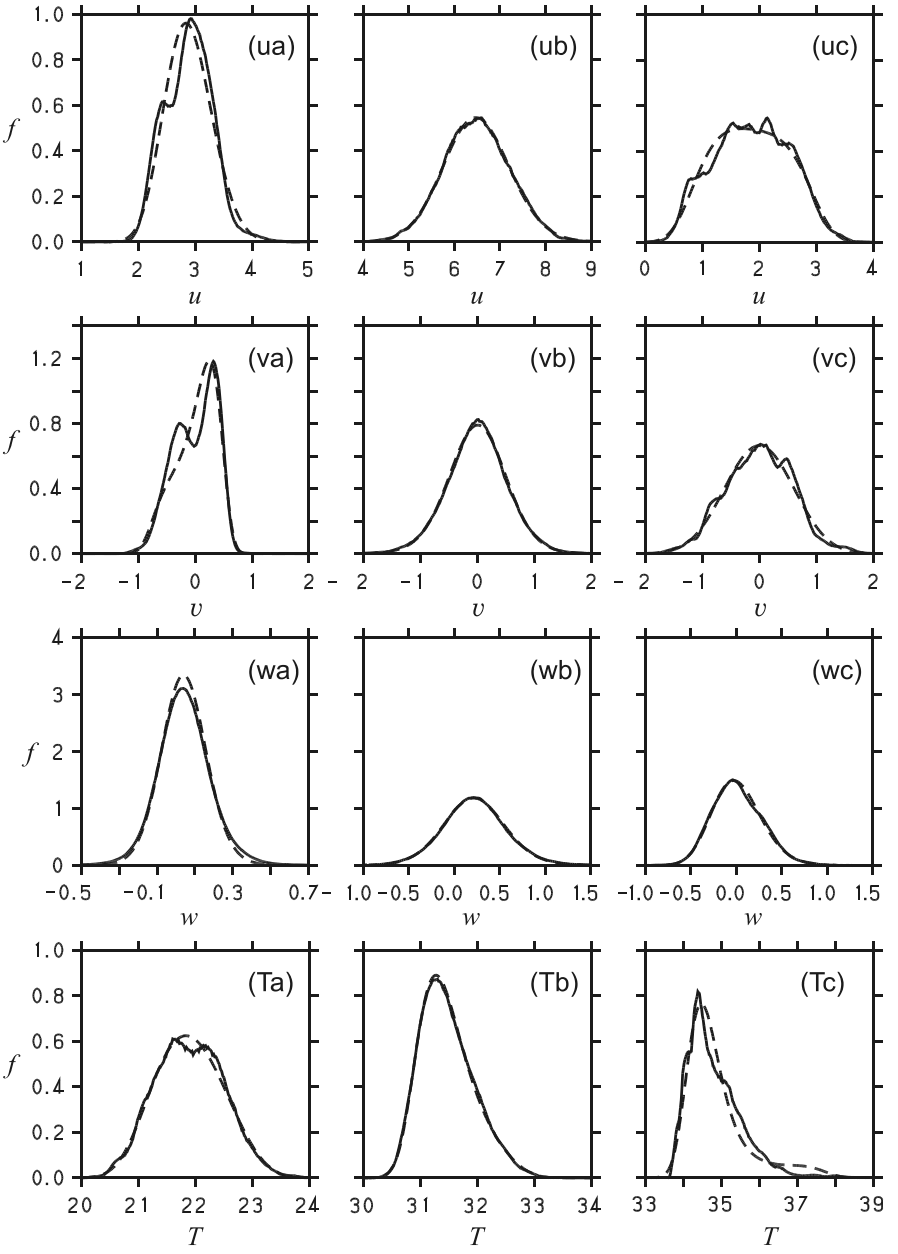



**Fig. 4.19.** The *solid lines* show velocity ($u$, $v$, $w$) and temperature ($T$) PDF data for stable (*left-hand side*), neutral (*middle*), and unstable (*right-hand side*) cases. The *dashed lines* show the fourth-order SML PDF model (4.132).

**Application of the Fourth-Order SML PDF Model.** The performance of the fourth-order SML PDF model (4.132) is shown in Fig. 4.19. The data required to determine the model parameters $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$ are given in Table 4.1, and $\lambda_0$ was calculated such that $f(x)$ integrates to one. Figure 4.19 shows that the fourth-order model (4.132) clearly improves the agreement between the modeled PDF and measured PDF. The model (4.132) is capable of accounting for the skewness (see the (Tb) and (Tc) figures) and the flatness of PDFs (see the (uc) figure). On the other hand, Eq. (4.132) is not a perfect model. The model performs better than the normal PDF model regarding the (ua) and (va) cases, but the model does not represent the bimodal mode structure of these PDFs. A way to improve the model performance would be given by extending the fourth-order PDF model to a sixth- (or higher) order PDF model. The obvious disadvantage of such an approach is the growing complexity of relations that determine the model parameters.

**Comparison with Other PDF Models.** The use of a SML PDF model is not the only way to address deficiencies of the normal PDF model. For example, a superposition of two normal PDFs can be used to model the PDFs shown in Fig. 4.17 (Luhar et al. 1996). However, it often turns out that the performance of such empirical PDF models is not better than the performance of the model (4.132). The significant disadvantage of such other models is related to the calculation of model parameters. For example, the PDF model involving two normal PDFs requires the numerical solution of two nonlinear equations to determine all the model parameters. The solution of these nonlinear equations is not simple: there may appear several solutions, or it is possible that there are no solutions. To avoid these numerical problems one can simplify these equations on the basis of empirical assumptions (Luhar et al. 1996). The disadvantage of this approach is that it is not easy to know under which conditions such empirical assumptions are applicable. These problems dot not appear if a SML PDF model is used. In this case, the model parameters are related analytically to data statistics.

## 4.6 Summary

Let us summarize the basic observations made in this chapter with regard to the analysis and modeling (in particular the availability and suitability of modeling concepts) of the stochastic state of one random variable.

**Data Analysis.** First of all, random data are analyzed in terms of their mean value and standard deviation. These quantities provide a basic understanding of the value and amount of randomness that we can expect. We are also interested in PDFs, which give us the possibility to calculate probabilities for certain events. A relevant feature of almost all real data is that there are usually a limited number of

observations. Such a limited number of samples has significant consequences for the calculation of mean values, standard deviations and PDFs: it means that these quantities do represent again random variables. PDFs are often also affected by the filtering applied to produce relatively smooth PDF shapes. In order to deal with this problem we need a basis that allows us to determine the standard of our results (we need to know whether a result obtained is just random, whether the result is close to the exact result, or whether we have an exact result). This basis is given by the first fundamental theorem of probability. This theorem allows us to determine the error due to a finite number of observations: we derived the relation $|\Delta_m| = |\Delta_{sd}| = |\varepsilon|\, N^{-1/2}$ for the relative errors of means and variances. A corresponding relation applies to PDFs, which do also represent a mean. Unfortunately, it needs huge numbers of samples to derive conclusions that are independent of the sample number. For example, we needs $10^4$ (or $10^6$) sample values to have an error of about 2% (or 0.2%).

**Modeling Concepts.** Which theoretical basis can we use to determine PDFs? There is only one general concept given by a SML PDF of a certain order. Such PDFs can be constructed such that any even number of PDF moments agrees with the same even number of moments of any data set. A normal PDF represents a second-order SML PDF, this means a normal PDF is the best PDF choice among all PDFs that agree with a given mean and standard deviation of data. Also, the second fundamental theorem of probability supports the use of a normal PDF under conditions where we do only have randomness (i.e., conditions where we do not have any deterministic processes in addition to randomness). For such conditions there is no need to use, for example, a fourth-order SML PDF model that accounts (in terms of the consideration of skewness and flatness) for deterministic processes in addition to randomness (see the discussion in Sect. 4.5). It is relevant to note that the concept of SML PDFs can be used for both unbounded and bounded variables (Heinz 2003). However, applications of SML PDFs to the case of bounded variables lead to nontrivial problems regarding the calculation of model parameters. For such cases it is often more convenient to use other PDFs, as given by gamma PDFs for non-negative random variables and beta PDFs for bounded variables. The advantage of gamma and beta PDFs is the simple relation between model parameters and means and variances. However, such PDFs do not have any theoretical support – they do only represent empirical assumptions. A significant disadvantage of such empirical concepts is that the extension of such concepts to the case of several variables is rather complicated due to the limited possibilities to account for correlations of random variables. The application of the SML PDF concept is also beneficial with regard to the latter question because SML PDFs can be constructed for several variables.

**Applicability of Modeling Concepts.** How useful are theoretical concepts for the modeling of PDFs? This question was addressed in terms of the discussion of the modeling of atmospheric velocity and temperature PDFs. It turns out that the normal PDF model represents a very valuable model: many PDFs can be accurately represented, and the normal PDF is at least a reasonable model in general. In particular, a normal PDF model is a very good model as long as only randomness is involved (as long as completely unorganized chaotic motions are considered). The normal PDF concept becomes questionable if deterministic trends are also involved (for example, oscillations of air parcels under stable conditions or upward and downward motions of warm and cold air parcels under unstable conditions). A way to improve the performance of models under conditions that involve deterministic trends is to develop SML PDFs of higher order, which usually improves the modeling performance. A disadvantage of such higher-order PDFs is that they need more information (e.g., knowledge about third, fourth, fifth and sixth-order central moments), and the complexity of model parameter calculations increases.

## 4.7 Exercises

**4.2.1** Consider $f(t) = \exp(-|t-t_0| / \tau) / (2\,\tau)$ with $\tau$ and $t_0$ as any model parameters.
a) Which properties of $f(t)$ are required conditions for the conclusion that $f(t)$ approaches a delta function, $f(t) \to \delta(t-t_0)$, for $\tau \to 0$?
b) Show analytically that $f(t) \to \delta(t-t_0)$ for $\tau \to 0$.

**4.2.2** Consider the uniform PDF given by Eq. (4.33).
a) Is it possible that the uniform PDF (4.33) behaves like a delta function? Explain the condition if the answer is yes.
b) If the answer is yes, is there (apart from the different variables applied) any difference between the delta function provided by the uniform PDF and the delta function provided by $f(t)$ considered in exercise 4.2.1?

**4.2.3** Show the correctness of the relation

$$\int g(x)\frac{d\delta(x-x_0)}{dx}dx = -\int \frac{dg}{dx}(x_0)\,\delta(x-x_0)\,dx = -\frac{dg}{dx}(x_0),$$

where $g(x)$ is any given function of $x$.

**4.2.4** A random number $X$ is uniformly distributed over the interval (–2, 3).
a) Calculate the probability $P(-3 \le X \le 0)$.
b) Calculate the probability $P(0 \le X)$.

**4.2.5** Consider the following problem (see Ross 2010). Buses arrive at a certain stop every 15 min starting at 7 a.m. A passenger arrives at the stop at a time that is uniformly distributed between 7:00 a.m. and 7:30 a.m. ($X$ refers to the number of minutes past 7 a.m., and $X$ is uniformly distributed over the interval (0, 30)).

a) Find the probability that he waits less than 2 min for a bus.

b) Find the probability that he waits more than 10 min for a bus.

**4.3.1** Consider the normal PDF (4.72) and the uniform PDF (4.33).

a) Calculate the entropy of the normal PDF.

b) Calculate the entropy of the uniform PDF.

c) Which PDF has the higher entropy if the normal PDF and uniform PDF have the same mean and variance?

d) Consider a normal PDF and uniform PDF that have the same mean and variance. The uniform PDF seems to represent the SML PDF: this PDF treats the probability for all possible states in the same way, there is no preference of certain states. Why is the normal PDF and not the uniform PDF the SML PDF?

**4.3.2** The intelligence quotient (IQ) is seen to be normally distributed with mean 100 and standard deviation 15.

a) What is the probability that the IQ of a randomly selected person will be between 80 and 120?

b) What is the probability that this IQ will be above 150?

**4.3.3** The heights of adult males in the United States are considered to be normally distributed with mean 1.75 m and standard deviation 0.07 m.

a) Calculate the probability that a randomly chosen adult male is between 1.5 m and 1.7 m tall.

b) What percentage of the adult male population is more than 1.95 m tall?

**4.3.4** The length of human gestation is seen to be normally distributed with mean 268 days and standard deviation 15 days.

a) What is the probability that a pregnancy takes 268 ± 3 days?

b) Consider this problem (see Ross 2010). The defendant in a suit is able to prove that he was out of the country during a period that began 290 days before the birth of the child and ended 240 days before the birth. If the defendant was the father of the child, what is the probability that the mother could have had the very long or very short gestation?

**4.3.5** The PDF $f(x)$ of a non-negative random variable $X$ has the structure

$$f(x) = \exp\left(a - b\,x - c\,x^2\right),$$

where $a$, $b$, and $c$ are any model parameters. It is assumed that $c \geq 0$ so that $f(x)$ does not diverge for large $x$ values.

a) Calculate the parameter $a$ in terms of the normalization condition.

b) Show that the mean value <$X$> implied by the PDF $f(x)$ is given by

$$\langle X \rangle = \frac{e^a - b}{2c}.$$

c) Show that the variance implied by this PDF is given by the expression

$$\left\langle \tilde{X}^2 \right\rangle = \frac{1 - b\langle X \rangle}{2c} - \langle X \rangle^2.$$

Hint: use the approach applied to the analysis of the normal PDF in Sect. 4.3.2 for doing this (see the second paragraph of this section).

**4.3.6** Consider again the PDF $f(x)$ given in exercise 4.3.5.

a) Specify the results a), b), and c) of exercise 4.3.5 for the case that $c = 0$. Hint: the simplest approach is to consider the PDF for $c = 0$ directly.

b) Explain the consequences of your results.

**4.3.7** Consider the PDF $f(x)$ and its consequences given in exercise 4.3.5.

a) Compare the PDF with other PDFs discussed in this chapter: which kind of PDF is considered?

b) Explain how the PDF parameters can be calculated so that the PDF $f(x)$ agrees with any given mean $\mu$ and any given variance $\sigma^2$.

**4.3.8** Consider the PDF $f(x)$ and its consequences given in exercise 4.3.5.

a) Develop a numerical scheme for finding the PDF parameters $a$, $b$, and $c$ so that the PDF agrees with any given mean $\mu$ and given variance $\sigma^2$.

b) Graph the PDF for the cases that $\sigma = 0.25$, $\sigma = 0.5$, and $\sigma = 1$, where $\mu = 1$ for all the three cases. Report the $b$ and $c$ values for these cases.

c) Discuss the three PDFs obtained in b) in comparison to the PDF shapes shown for the gamma PDF in Fig. 4.9.

**4.3.9** Consider the gamma PDF discussed in Sect. 4.3.3.

a) Calculate the skewness $m_3$ and flatness $m_4$ for the gamma function.

b) Under which condition are the skewness and flatness values equal to the values $m_3 = 0$ and $m_4 = 3$ of a normal PDF?

c) What does this condition mean regarding the mean value and variance?

d) Which PDF shape is related to this condition?

**4.3.10** Consider the beta PDF discussed in Sect. 4.3.4.

a) Show that the moments of the beta PDF ($k = 1, 2, \ldots$) are given by

$$\left\langle X^k \right\rangle = \frac{(a+k-1)(a+k-2)\cdots a}{(a+b+k-1)(a+b+k-2)\cdots(a+b)}.$$

b) Use these moment relations to show that the PDF parameters $a$ and $b$ are related to the mean and variance via the relations

$$a = \langle X \rangle \left[ \frac{\langle X \rangle (1 - \langle X \rangle)}{\langle \widetilde{X}^2 \rangle} - 1 \right], \qquad b = (1 - \langle X \rangle) \left[ \frac{\langle X \rangle (1 - \langle X \rangle)}{\langle \widetilde{X}^2 \rangle} - 1 \right].$$

**4.3.11** Consider the beta PDF discussed in Sect. 4.3.4. The skewness $m_3$ and the flatness $m_4$ of the beta function are given by the expressions

$$m_3 = 2\frac{b-a}{a+b+2}\sqrt{\frac{a+b+1}{ab}}, \qquad m_4 = 3\frac{a+b+1}{a+b+3}\left(1 + \frac{2(b-a)^2}{ab(a+b+2)}\right).$$

a) Under which condition are the skewness and flatness values equal to the values $m_3 = 0$ and $m_4 = 3$ of a normal PDF?
b) What does this condition mean regarding the mean value and variance?
c) Which PDF shape is related to this condition?

**4.3.12** Consider the following problem (Stewart 2006). The hydrogen atom consists of one proton in the nucleus and one electron, which moves about the nucleus. The electron does not move in a well-defined orbit, but there is a probability for finding the electron at a certain distance from the nucleus. The PDF is given by $p(r) = 4\, r^2 \exp(-2\, r / a_0) / a_0^3$ for $r \geq 0$, where $a_0 = 5.59 \times 10^{-11}$ m is the Bohr radius. The integral over this PDF from zero to $r$ gives the probability that the electron will be found within a sphere of radius $r$ meters centered at the nucleus.
a) Calculate the mean distance $R$ of the electron from the nucleus.
b) Calculate the variance $\sigma^2$ of the PDF.
c) Show that the PDF is a gamma PDF. Hint: use $\Gamma(n+1) = n!$ for integer values of $n$.
d) For what value of $r$ does $p(r)$ have its maximum value?
e) Find the probability that the electron will be within the sphere of radius $3\, a_0$ centered at the nucleus.
Hint: the following integrals will be helpful ($a$ is any constant).

$$\int x^2 e^{ax} dx = \frac{e^{ax}}{a^3}\left(a^2x^2 - 2ax + 2\right), \qquad \int x^3 e^{ax} dx = \frac{e^{ax}}{a^4}\left(a^3x^3 - 3a^2x^2 + 6ax - 6\right),$$

$$\int x^4 e^{ax} dx = \frac{e^{ax}}{a^5}\left(a^4x^4 - 4a^3x^3 + 12a^2x^2 - 24ax + 24\right).$$

**4.4.1** The variance $\sigma^2_{Mean}$ of the mean $\mu_N$ of $N$ independent samples is defined by

$$\sigma^2{}_{Mean} = \left\langle \left[\mu_N - \langle X \rangle\right]^2 \right\rangle = \left\langle \left[ \frac{1}{N} \sum_{i=1}^{N} X_i - \langle X \rangle \right]^2 \right\rangle.$$

a) Show that $\sigma^2_{Mean} = \sigma^2 / N$, where $\sigma^2 = \langle (X - \langle X \rangle)^2 \rangle$.
b) Relate $\sigma_{Mean} = \sigma / N^{1/2}$ obtained here to the finding $|\Delta_m| = |\Delta_{sd}| = |\varepsilon| / N^{1/2}$ obtained in Sect. 4.4.2.

**4.4.2** Consider the variance $\sigma^2_N$ of $N$ independent samples, which is defined by

$$\sigma^2{}_N = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \mu_N\right)^2.$$

a) Calculate the mean value $\langle \sigma^2_N \rangle$ to show that this expression represents an unbiased estimate for the variance (to show that $\langle \sigma^2_N \rangle = \sigma^2$, where $\sigma^2 = \langle (X - \langle X \rangle)^2 \rangle$). Hint: Replace $X_i - \mu_N$ in $\sigma^2_N$ by $X_i - \langle X \rangle - (\mu_N - \langle X \rangle)$. Distribute the square. Use the definitions of $\sigma^2$ and $\sigma^2_{Mean} = \sigma^2 / N$.
b) How relevant is Bessel's correction (the use of $N - 1$ instead of $N$ in the variance formula)?

**4.4.3** Consider the case that there are insufficient data to estimate a meaningful variance. What is the best possible characterization of any random variable $X$ for this case? Hint: consider the mean squared error $g(x) = \langle (X - x)^2 \rangle$ of any estimate $x$ of $X$ and calculate the minimum of $g(x)$.

**4.5.1** Consider Eqs. (4.136) for the calculation of the model parameters of the fourth-order SML PDF model (see Sect. 4.5.3).
a) Use Eqs. (4.136) for calculating the model parameters of a second-order SML PDF model.
b) For which data set do these relations imply that the fourth-order SML PDF reduces to a second-order SML PDF model?

**4.5.2** Consider the fourth-order SML PDF model discussed in Sect. 4.5.3. The relations (4.136) represent four equations for the four unknowns $\lambda_1$, $\lambda_2$, $\lambda_3$, and $\lambda_4$. According to the model development, $\lambda_1$, $\lambda_2$, $\lambda_3$, and $\lambda_4$ can be calculated by knowing the mean, the variance, $m_3$ and $m_4$.
a) Explain how the parameters $m_5$ and $m_6$ are mathematically determined in terms of Eqs. (4.136).
b) Explain the use of Eqs. (4.136) for the calculation of a fourth-order SML PDF without knowing anything about $m_5$ and $m_6$.

**4.5.3** Consider the temperature anomaly data presented in Table 1.7.
a) Consider the data from 1850–1930. Calculate the filtered PDF for these data for three choices of the filter interval $\Delta T$. Which $\Delta T$ represents the best choice?
b) Consider the data from 1930–2008. Calculate the filtered PDF for these data by using the best available $\Delta T$ determined in a).
c) Use the comparison of these two PDFs to predict the potential features of the PDF for the next eighty years.

# 5 Deterministic Changes

The models discussed in the previous chapters were based on observations: the goal was to develop models that agree with any given observations. A different modeling approach is the use of mathematical concepts that explain changes of variables (e.g., by means of a formula that provides the change of any population over a certain time). The comparison of predictions implied by such concepts with observations is usually needed to show the validity of the ideas applied, but the model development can be performed without the use of observations. Usually, such model developments are performed on the basis of difference equations or differential equations. Differential equations are applied for postulating general laws (e.g., the laws of mechanics or the laws of population ecology). Characteristic features of such equations for the deterministic evolution of single variables and several variables will be described in Chaps. 7 and 9, respectively. Difference equations are useful for the modeling of processes that require the modeling of changes over a specific time period (e.g., the change of any population per year). Apart from that, difference equations provide the basis for the mathematical formulation, numerical solution, and in some cases even the analytical solution of differential equations. Therefore, characteristic features of deterministic difference equations will be discussed in this chapter in order to prepare the discussion of differential equations in Chaps. 7 and 9.

Section 5.1 explains the motivation for this modeling approach. The simplest way to use difference equations for the modeling of deterministic changes (linear first-order difference equations) will be discussed in Sect. 5.2. Sections 5.3 and 5.4 describe extensions of such equations: characteristic features of second-order linear equations and first-order nonlinear difference equations will be discussed. Relations between difference and differential equations and the basic advantages and disadvantages of these equations will be described in Sect. 5.5. The observations on difference equation models will be summarized in Sect. 5.6.

S. Heinz, *Mathematical Modeling*, DOI 10.1007/978-3-642-20311-4_5,

## 5.1 Motivation

**National Income.** Let us discuss an example to illustrate the benefits of using difference equations for the modeling of changes. The national income represents a measure for the total level of economic activity within a country during a certain period of time. The economic activity of a country determines the living standard of people: a recession decreases and a boom increases the living standard. Thus, there is a great interest in the variations of the national income to understand, for example, the effect of factors that influence the national income. There is a variety of possibilities to define the national income and its influence factors. Here, we will focus on a relatively simple model pioneered by the American economist P. A. Samuelson (1939). Samuelson's model describes the national income $y_n$ of a country during an accounting period $n = 0, 1, 2, \ldots$ (n may refer to one year) as a sum of three contributions,

$$y_n = C_n + I_n + G_n. \tag{5.1}$$

$C_n$ is the amount spent on consumer goods (food, housing, clothing, …) during the accounting period. $I_n$ is the induced private investment (the amount invested in new machinery, equipment, …). $G_n$ is the amount spent by the government (social security, health, education, infrastructure investments, research spending, …). To calculate $y_n$ it is needed to relate $C_n$, $I_n$, and $G_n$ to $y_n$. The consumption expenditure $C_n$ may be expected to be proportional to the national income of the preceding period (an increase in the living standard leads to an increase in consumption),

$$C_n = \alpha\, y_{n-1}. \tag{5.2}$$

Here, the constant of proportionality $0 \le \alpha \le 1$ is called the marginal propensity to consume. The induced private investment $I_n$ can be modeled in several ways. One possible assumption (which is called the acceleration principle) is that $I_n$ is proportional to the increase in consumption of the period considered over the preceding period (an increasing living standard over the previous period leads to increased investments to offer additional goods),

$$I_n = \beta(C_n - C_{n-1}) = \alpha\beta(y_{n-1} - y_{n-2}), \tag{5.3}$$

where the model parameter $\beta \ge 0$. Samuelson considered variations $0 \le \beta \le 4$. Finally, a simple model for the government expenditure $G_n$ is that $G_n$ is constant,

$$G_n = \gamma. \tag{5.4}$$

By using appropriate units we may assume that $\gamma = 1$. According to the three assumptions (5.2)–(5.4), Samuelson's national income model reads

$$y_n = \alpha\, y_{n-1} + \alpha\beta(y_{n-1} - y_{n-2}) + \gamma = \alpha(1+\beta)\, y_{n-1} - \alpha\beta\, y_{n-2} + \gamma. \tag{5.5}$$

Starting with any initial value $y_0$, the model (5.5) can be used to calculate the evolution of the national income in dependence on $y_0$ and the model parameters $\alpha$, $\beta$, and $\gamma$. This model is capable of describing a wide range of scenarios. A detailed analysis of features of this model will be provided in Sect. 5.3.

**Questions About Difference Equation Models.** The national income problem considered in the preceding paragraph (or any other problem involving difference equations) leads to several questions related to the use of such equations for the modeling of problems:

- Which processes can be described by which type of difference equation?
- How can we solve a certain difference equation analytically?
- How is it possible to adjust model parameters to given observations?

As mentioned in the beginning, there is a close relationship between difference and differential equations. With regard to these relations, there are questions like:

- How are difference and differential equations related to each other?
- What is the advantage and disadvantage of using difference equation models?
- What is the advantage and disadvantage of using differential equation models?

These and other questions will be addressed in the following by analyzing the features of difference equations and illustrating the benefits of such models by means of applications.

## 5.2 Linear Changes

Let us discuss first the simplest concept for a linear difference equation: a first-order linear difference equation for $y_n$. We assume that $y_n$ has a given initial value $y_0$ at $n = 0$. The $y_n$ values for $n = 1, 2, \ldots$ are determined by the equation

$$y_n = a\, y_{n-1} + b. \tag{5.6}$$

Here, $a$ and $b$ are any model parameters. This equation can be written $y_n - y_{n-1} = (a-1)\, y_{n-1} + b$, i.e., Eq. (5.6) provides a formula for changes of $y_n$. Equation (5.6) is a linear equation because $y_n$ does only appear linearly. This equation is called a first-order equation because only $y_n$ and $y_{n-1}$ are involved (i.e., there are no terms $y_{n-2}$, or $y_{n-3}$, …). Equation (5.6) is called a homogeneous equation if $b = 0$.

### *5.2.1 Linear First-Order Difference Equations*

**Solution.** Let us try to find the solution to Eq. (5.6). The solution is a formula that gives us the value of $y_n$ based on the knowledge of the parameters $a$, $b$, and $y_0$ involved. To find the solution we have to understand how the value of $y_n$ is

affected by the initial value $y_0$. To see this we consider $y_n$ for $n = 1, 2, \ldots$,

$$y_1 = a\,y_0 + b, \tag{5.7a}$$

$$y_2 = a\,y_1 + b, \tag{5.7b}$$

$$y_3 = a\,y_2 + b. \tag{5.7c}$$

To relate $y_n$ to $y_0$, we use $y_1$ in the relation for $y_2$, and $y_2$ in the expression for $y_3$,

$$y_1 = a\,y_0 + b, \tag{5.8a}$$

$$y_2 = a\big(a\,y_0 + b\big) + b = a^2 y_0 + b(a+1), \tag{5.8b}$$

$$y_3 = a\Big(a\big(a\,y_0 + b\big) + b\Big) + b = a^3 y_0 + b(a^2 + a + 1). \tag{5.8c}$$

According to these expressions, it is reasonable to expect that the general solution has the form (n = 0, 1, …),

$$y_n = a^n y_0 + b(a^{n-1} + a^{n-2} + \cdots + a^1 + a^0). \tag{5.9}$$

**Solution Rewriting.** The sum of powers of $a$ represents a geometric series,

$$a^{n-1} + a^{n-2} + \cdots + a^1 + a^0 = \frac{1-a^n}{1-a}. \tag{5.10}$$

The validity of Eq. (5.10) can be seen by introducing the partial sum

$$p_n = a^{n-1} + a^{n-2} + \cdots + a^1 + a^0. \tag{5.11}$$

The multiplication of this relation with $a$ results in

$$a\,p_n = a^n + a^{n-1} + a^{n-2} + \cdots + a^1. \tag{5.12}$$

By taking the difference $p_n - a\,p_n$ we find that all terms cancel with the exception of two terms:

$$p_n - a\,p_n = 1 - a^n. \tag{5.13}$$

The latter relation then provides

$$p_n = \frac{1-a^n}{1-a}, \tag{5.14}$$

which shows the validity of Eq. (5.10). The combination of Eqs. (5.9) and (5.10) enables us to write the solution in the following way ($n = 0, 1, \ldots$),

$$y_n = a^n\,y_0 + b\frac{1-a^n}{1-a} = \frac{b}{1-a} + a^n\left(y_0 - \frac{b}{1-a}\right). \tag{5.15}$$

The consistency of this solution at $a = 1$ can be seen by using Eq. (5.9), or by using l'Hospital's rule. In both ways we find for $a = 1$ that

$$y_n = y_0 + nb. \tag{5.16}$$

**Solution Proof.** To prove the correctness of our assumption that the solution $y_n$ is given by Eq. (5.15) it is needed to show first that this solution is in agreement with the initial value, and second that this solution satisfies the difference equation (5.6). By setting $n = 0$ on both sides of Eq. (5.15) we find

$$y_0 = y_0, \tag{5.17}$$

which means that the solution is in agreement with the initial value. Evidence that the solution satisfies the difference equation can be obtained by using Eq. (5.15) in the difference equation (5.6),

$$\begin{aligned} &\frac{b}{1-a} + a^n\left(y_0 - \frac{b}{1-a}\right) = a\left[\frac{b}{1-a} + a^{n-1}\left(y_0 - \frac{b}{1-a}\right)\right] + b \\ &\quad = b\left(1 + \frac{a}{1-a}\right) + a^n\left(y_0 - \frac{b}{1-a}\right) = \frac{b}{1-a} + a^n\left(y_0 - \frac{b}{1-a}\right). \end{aligned} \tag{5.18}$$

The rewriting of the last expression in the first line results in the same expression on both sides. Thus, that the solution (5.15) solves the difference equation (5.6).

**Solution Features.** Characteristic features of solutions to the first-order linear difference equation (5.6) are illustrated in Fig. 5.1. Fixed values $b = 1$ and $y_0 = 0.1$ are applied, and a is given by the values shown in Fig. 5.1. These figures show that the value of $y_n$ becomes constant for large $n$ if $|a| < 1$. The equilibrium value s is given by

$$s = \frac{b}{1-a}, \tag{5.19}$$

as may be seen from Eq. (5.15). The value $a = 1$ implies a linear function according to Eq. (5.16), and values $a > 1$ lead to increasing functions. Fluctuations of $y_n$ appear if the values of a are negative. The fluctuations are damped out, or they have a constant or an increasing amplitude. With regard to applications, it turns out that models that involve such deterministic fluctuations are not often helpful – it is usually hard to adjust such models to data.

### *5.2.2 Interest*

**Interest.** Let us consider some applications of first-order difference equations to finance problems (Fulford et al. 1997) to illustrate the benefits of using such equations. A sum of money that is lent to a bank earns interest. Let us denote the amount on deposit after $n$ time periods by $S_n$. The interest is denoted by $I_n$. The relation between $S_n$ and $I_n$ is given by

$$S_n = S_{n-1} + I_{n-1}. \tag{5.20}$$

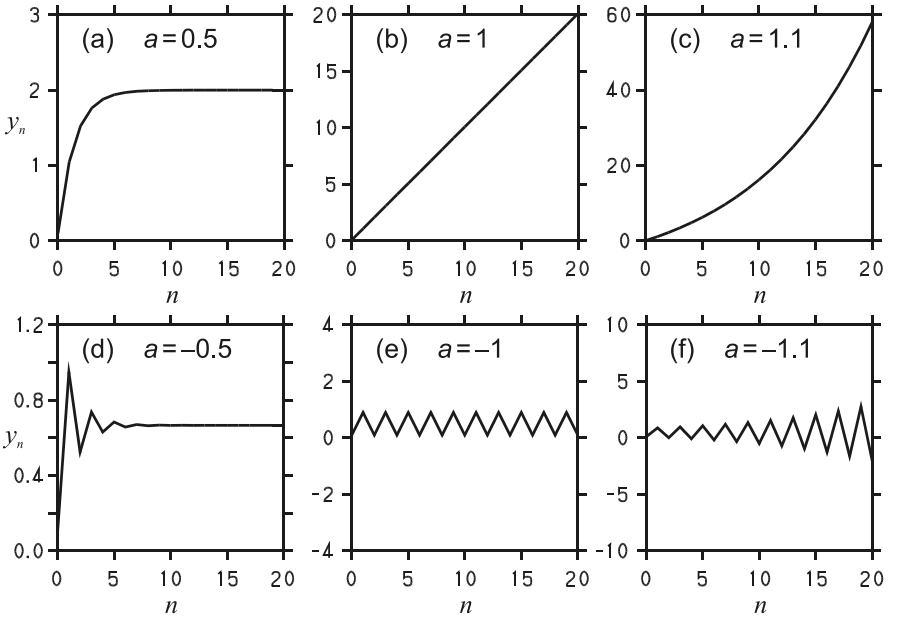



**Fig. 5.1.** An illustration of features of solutions of the first-order linear difference equation (5.6) in dependence on $n$, where $b = 1$ and $y_0 = 0.1$. The values of the model parameter $a$ are shown in the plots.

This formula states that the amount on deposit after $n$ periods is equal to the sum of the amount on deposit after the previous period plus the interest earned. The question now is how the interest $I_n$ is related to $S_{n-1}$.

**Simple Interest.** A simple way of offering interest is given by the option that the interest is proportional to the initial sum $S_0$ deposited, called the principal,

$$I_{n-1} = p\,S_0. \tag{5.21}$$

Here, p is the constant interest rate (for example, $p = 0.05$ means that there is an interest rate of 5%). The use of this expression in (5.20) results in the following first-order difference equation for $S_n$ ($n = 0, 1, \ldots$),

$$S_n = S_{n-1} + p\,S_0. \tag{5.22}$$

This equation represents a specific case of Eq. (5.6) where $a = 1$ and $b = p\,S_0$. By using the solution (5.16) for $a = 1$, the solution of Eq. (5.22) is given by

$$S_n = S_0 + n\,p\,S_0 = (1 + n\,p)\,S_0. \tag{5.23}$$

This formula is called the simple interest formula. According to this option, the amount on deposit after $n$ periods is a linear function of the principal $S_0$. It is assumed that $n$ refers to the number of years.

**Example.** An example for the use of the simple interest option is the following one. We are interested to know the number $k$ of years needed to double any initial interest $S_0$. According to (5.23), the condition for the calculation of $k$ is given by

$$2S_0 = (1 + k\,p)S_0. \tag{5.24}$$

Therefore, the number $k$ of years required is $k = 1 / p$. For example, 50 years are required to double any principal if the interest rate is $p = 0.02$.

**Compound Interest.** Another type of interest is given if interest is added to the principal at regular intervals, which are called conversion periods. For this case, the interest formula is given by

$$I_{n-1} = \frac{p}{r} S_{n-1}. \tag{5.25}$$

Here, $r$ refers to the number of conversion periods per year, this means $r = 12$ corresponds to a conversion period of one month, and $r = 4$ corresponds to a conversion period of one quarter. For this interest option, the difference equation for $S_n$ is given by ($n = 0, 1, \ldots$),

$$S_n = S_{n-1} + \frac{p}{r} S_{n-1} = \left(1 + \frac{p}{r}\right) S_{n-1}. \tag{5.26}$$

This equation is an homogeneous first-order linear difference equation. To use the solution formula (5.15), it is required to relate the model parameters $p$ and $r$ to the parameters of the general difference equation (5.6). This comparison shows that $a = 1 + p / r$ and $b = 0$. Using $a$ and $b$ from Eq. (5.15) in Eq. (5.26) yields

$$S_n = \left(1 + \frac{p}{r}\right)^n S_0. \tag{5.27}$$

This formula is called the compound interest formula. Regarding the use of this formula it is relevant to note that $n$ refers to the number of conversion periods (rather than the number of years in the simple interest formula (5.23)). For example, to find the amount of deposit after 10 years at compound interest with a monthly conversion period, one has to apply $n = 120$ in Eq. (5.27).

**Effect of Conversion Period.** What will be the effect of different conversion periods? To address this question, let us consider the condition for the validity of

$$\left(1 + \frac{p}{r_1}\right)^{r_1 n} \geq \left(1 + \frac{p}{r_2}\right)^{r_2 n}, \tag{5.28}$$

which compares the effect of two conversion periods on the interest. We take the natural logarithm and divide both sides by $n\,p$. This leads to

$$\frac{r_1}{p} \ln\left(1 + \frac{p}{r_1}\right) \geq \frac{r_2}{p} \ln\left(1 + \frac{p}{r_2}\right). \tag{5.29}$$

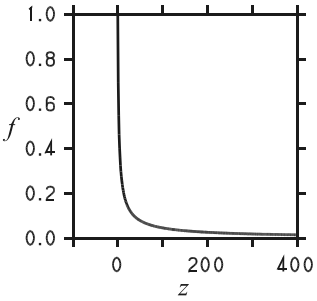



**Fig. 5.2.** The function $f(z)$ according to Eq. (5.30).

By introducing the function

$$f(z) = \frac{\ln(1+z)}{z}, \tag{5.30}$$

Eq. (5.29) can be also written

$$f(z_1) \geq f(z_2), \tag{5.31}$$

where $z_1 = p / r_1$ and $z_2 = p / r_2$. Figure 5.2 shows that $f(z)$ is a decreasing function of $z$. Therefore, the condition for the validity of Eq. (5.28) is that $z_1 \leq z_2$, which requires that $r_1 \geq r_2$. Thus, the higher the number of conversion periods, the higher will be the interest made over the same time at a constant interest rate $p$.

**Comparison of Options.** Let us compare the compound interest and the simple interest options. This comparison of options can be performed by making use of Bernoulli's inequality

$$(1+z)^n \geq 1 + nz, \tag{5.32}$$

where $z \geq 0$ and the integer $n \geq 0$. The validity of Bernoulli's inequality can be seen by using the binomial theorem for the left-hand side,

$$(1+z)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} z^k = 1 + n\,z + \Delta_n, \tag{5.33}$$

where $\Delta_n$ is a non-negative contribution. We may set $z = p / r$ and replace $n$ by $r\,n$ in Bernoulli's inequality,

$$\left(1 + \frac{p}{r}\right)^{r\,n} \geq 1 + r\,n\,\frac{p}{r} = 1 + n\,p. \tag{5.34}$$

Here, $n$ refers to the number of years considered. The right-hand side multiplied with $S_0$ gives the amount on deposit made with the simple interest option. The left-hand side multiplied with $S_0$ gives the amount on deposit made over the same time (over $r\,n$ conversion periods) with the compound interest option. Therefore, the compound interest option produces over the same amount of time at least as much interest as the simple interest option if $S_0$ and $p$ are the same.

**Example.** To illustrate the use of the compound interest formula we consider the same example as before: we calculate the number $k$ of years needed to double any interest $S_0$. According to Eq. (5.27), the condition for $k$ is now given by

$$2S_0 = \left(1 + \frac{p}{r}\right)^k S_0. \tag{5.35}$$

By taking the natural logarithm of both sides, we find the number $k$ of years to be determined by the relation

$$\ln 2 = k \ln\left(1 + \frac{p}{r}\right). \tag{5.36}$$

Thus, it takes 416.2 conversion periods (34.7 years) to double any principal if the interest rate is $p = 0.02$ and the interest is compounded monthly. The compound interest produces the same amount on deposit as the simple interest in less time, which agrees with the conclusion obtained in the preceding paragraph.

### *5.2.3 Loan Repayments*

**Loan Repayments.** Loan repayments can be studied in a corresponding way. Let us consider the repayment of a housing loan that required any initial debt $D_0$ (the amount borrowed from a financial institution). The debt changes due to (*i*) repayments at regular intervals made to reduce the debt, and (*ii*) interest that has to be paid on the amount still owing. By considering a compound interest, the debt $D_n$ after $n$ payments is given by the following formula

$$D_n = \left(1 + \frac{p}{r}\right) D_{n-1} - R. \tag{5.37}$$

Here, $D_n$ and $D_{n-1}$ are the debt after $n$ and $(n-1)$ payments, respectively, and $R$ is the constant repayment. The term involving $p / r$ refers to the compound interest option. To enable the use of the general solution formula (5.15) it is needed to relate the parameters of the model (5.37) to the parameters $a$ and $b$ in the general difference equation (5.6). This comparison reveals that $a = 1 + p / r$ and $b = -R$. According to Eq. (5.15), the solution $S_n$ of Eq. (5.37) is, therefore, given by

$$D_n = \left(1 + \frac{p}{r}\right)^n D_0 - R\frac{1 - (1 + p/r)^n}{1 - (1 + p/r)} = \left(1 + \frac{p}{r}\right)^n D_0 + \frac{rR}{p}\left[1 - \left(1 + \frac{p}{r}\right)^n\right]. \tag{5.38}$$

This formula is called the loan repayment formula. In correspondence to the compound interest formula (5.27), $n$ refers to the number of conversion periods.

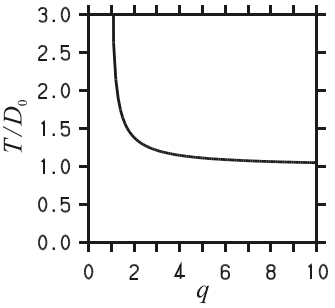



**Fig. 5.3.** $T/D_0$ as a function of $q$ according to Eq. (5.44).

**Example.** To illustrate the use of Eq. (5.38), let us find first the repayment $R$ needed to keep the debt equal to the initial debt $D_0$. By setting $D_n = D_0$, Eq. (5.38) can be written

$$D_0\left[1-\left(1+\frac{p}{r}\right)^n\right]=\frac{rR}{p}\left[1-\left(1+\frac{p}{r}\right)^n\right]. \tag{5.39}$$

The bracket terms are nonzero for nonzero $p$ and $r$ and $n > 0$. Hence, this relation implies $R = p D_0 / r$. Let us assume now that a monthly repayment of

$$R = q\frac{p}{r}D_0 \tag{5.40}$$

is made to reduce the debt, where $q \geq 1$. The number $k$ of conversion periods required to repay the debt can be calculated for this case on the basis of Eq. (5.38). By using $R = q p D_0 / r$ and $D_0 = 0$ in Eq. (5.38), we obtain for $k$ the condition

$$0=\left(1+\frac{p}{r}\right)^k D_0+qD_0\left[1-\left(1+\frac{p}{r}\right)^k\right]=D_0\left\{q-(q-1)\left(1+\frac{p}{r}\right)^k\right\}. \tag{5.41}$$

The corresponding condition for $k$ is given by

$$k=\ln\left(\frac{q}{q-1}\right)\left[\ln\left(1+\frac{p}{r}\right)\right]^{-1}. \tag{5.42}$$

The total payment $T$ made over $k$ conversion periods is $k$ times $R = q p D_0 / r$, i.e.,

$$T=\frac{q\,p\,D_0}{r}\ln\left(\frac{q}{q-1}\right)\left[\ln\left(1+\frac{p}{r}\right)\right]^{-1}. \tag{5.43}$$

For $p / r << 1$ (as given for example for the case that the interest rate is $p = 0.02$ and the interest is compounded monthly), $\ln(1 + p / r) = p / r$ represents a very good approximation. For this case, Eq. (5.43) can be written

$$T = q\ln\left(\frac{q}{q-1}\right)D_0. \tag{5.44}$$

The variation of $T$ standardized by $D_0$ as function of $q$ is shown in Fig. 5.3. It may be seen that monthly payments of $R = 2\,p\,D_0\,/\,r$ are a requirement to ensure that the total payment is below 1.4 $D_0$. It is interesting to observe that higher payments than $R = 2\,p\,D_0\,/\,r$ have a rather limited effect on the total costs.

## 5.3 Linear Changes with Delays

The second type of difference equation considered is a second-order linear difference equation. Initial values $y_0$ at $n = 0$ and $y_1$ at $n = 1$ are considered to be given. The $y_n$ values for n = 2, 3, … are determined by the equation

$$y_n = A\,y_{n-1} + B\,y_{n-2} + C, \tag{5.45}$$

where $A$, $B$, and $C$ are any model parameters. By adding $-\,y_{n-1}$ on both sides, this equation can be also written as an equation for changes $y_n - y_{n-1}$. Equation (5.45) is called homogeneous if $C = 0$. Equation (5.45) is linear because there are only terms that are linear in the function considered. Equation (5.45) is a second-order difference equation due to the inclusion of $y_{n-2}$ in addition to $y_n$ and $y_{n-1}$. The consideration of $y_{n-2}$ is the reason for talking about the inclusion of delay effects. The concept of a first-order difference equation is that the current state $y_{n-1}$ determines the future state $y_n$. A second-order difference equation extends the concept of a first-order equation by accounting also for an effect of the previous state $y_{n-2}$ on the future state $y_n$. The latter effect can be seen to be delayed.

### *5.3.1 Linear Second-Order Difference Equations*

**Solution Approach.** It turns out that the simple approach used to solve a first-order linear equation is inappropriate to find a solution for the second-order linear equation (5.45). One may expect that the solution will involve two contributions: one equilibrium solution (denoted by $s$), and one transitional solution (denoted by $z_n$) that accounts for the transition to the equilibrium. Hence, we write

$$y_n = s + z_n. \tag{5.46}$$

The equilibrium solution $s$ can be found by replacing $y_n$ by $s$ in Eq. (5.45),

$$s = A\,s + B\,s + C. \tag{5.47}$$

This relation determines the equilibrium solution as

$$s = \frac{C}{1 - A - B}. \tag{5.48}$$

It is worth noting that such a constant equilibrium solution does not always exist: see exercise 5.3.3. The $z_n$ equation can be found by using $y_n = s + z_n$ in Eq. (5.45),

$$\frac{C}{1-A-B} + z_n = A\left(\frac{C}{1-A-B} + z_{n-1}\right) + B\left(\frac{C}{1-A-B} + z_{n-2}\right) + C, \tag{5.49}$$

where $s = C/(1-A-B)$ is applied. This equation can be written

$$z_n - A z_{n-1} - B z_{n-2} = \frac{C}{1-A-B}\left(1-A-B-1+A+B\right) = 0. \tag{5.50}$$

Hence, the equation for the transitional solution $z_n$ reads

$$z_n = A z_{n-1} + B z_{n-2}. \tag{5.51}$$

This is an homogeneous equation as there is no term independent of the function $z_n$ for which we solve. To determine the structure of solutions to Eq. (5.51), let us consider solutions for the case that $A = 0$ and $B = 0$, respectively. For $B = 0$ we find the solution $z_n = A^n z_0$, which corresponds to the solution of the linear first-order equation (5.6) for the case that $b = 0$. For the case $A = 0$ we find the solution $z_n = \pm B^{n/2} z_0$. The solution for the latter case can be found by assuming a similar form of the solution as given for the $B = 0$ case. The validity of $z_n = A^n z_0$ and $z_n = \pm B^{n/2} z_0$ can be seen by using these expressions in (5.51) for the cases considered. The solutions obtained lead to two conclusions. First, there are two solutions of Eq. (5.51) in general. Second, we may assume that each solution consists of a constant multiplied with another raised to the power $n$. Therefore, solutions to the homogeneous equation (5.51) can be expected to have the structure

$$z_n = c_1 r_1^{\,n} + c_2 r_2^{\,n}, \tag{5.52}$$

where $c_1$, $c_2$, $r_1$, and $r_2$ are any constants that have to be specified. According to $y_n = s + z_n$, the general solution to Eq. (5.45) is expected to be given by

$$y_n = s + c_1 r_1^{\,n} + c_2 r_2^{\,n}. \tag{5.53}$$

To demonstrate the validity of this assumption we have to show first that this expression satisfies the difference equation (5.45). This demonstration will show how $r_1$ and $r_2$ are related to the model parameters $A$ and $B$. Second, we have to explain under which condition this expression agrees with the initial data. The latter will show how $c_1$ and $c_2$ are related to $y_0$ and $y_1$.

**Consistency with Difference Equation.** To prove the suitability of Eq. (5.53), we have to show that this expression can satisfy the difference equation (5.45). According to the discussion in the preceding paragraph, the latter requires evidence that the assumption (5.52) for $z_n$ can satisfy the $z_n$ equation (5.51), i.e., we have to show that the following relation can be satisfied,

$$c_1 r_1^{\,n} + c_2 r_2^{\,n} = A\left(c_1 r_1^{\,n-1} + c_2 r_2^{\,n-1}\right) + B\left(c_1 r_1^{\,n-2} + c_2 r_2^{\,n-2}\right). \tag{5.54}$$

The latter condition can be also written

$$c_1 r_1^{n-2}\left(r_1^2 - A r_1 - B\right) + c_2 r_2^{n-2}\left(r_2^2 - A r_2 - B\right) = 0. \tag{5.55}$$

The terms $c_1 r_1^{n-2}$ and $c_1 r_1^{n-2}$ have to be nonzero to enable the calculation of non-trivial solutions. Thus, Eq. (5.55) requires that the bracket terms are zero,

$$r_1^2 - A r_1 - B = 0, \qquad r_2^2 - A r_2 - B = 0. \tag{5.56}$$

The conditions (5.56) represent the same quadratic equation for the two roots $r_1$ and $r_2$. Thus, $r_1$ and $r_2$ can be calculated as roots of the equation

$$r^2 - A r - B = 0. \tag{5.57}$$

This equation is called the characteristic equation. It determines two solutions

$$r_1 = \frac{A}{2} + \sqrt{\frac{A^2}{4} + B}, \qquad r_2 = \frac{A}{2} - \sqrt{\frac{A^2}{4} + B}. \tag{5.58}$$

The two roots $r_1$ and $r_2$ are the eigenvalues of the characteristic equation (5.57). To prepare the following developments we write $r_1$ and $r_2$ as

$$r_1 = r_S + r_D, \qquad r_2 = r_S - r_D, \tag{5.59}$$

where $r_S$ and $r_D$ are given by

$$r_S = \frac{r_1 + r_2}{2} = \frac{A}{2}, \qquad r_D = \frac{r_1 - r_2}{2} = \sqrt{\frac{A^2}{4} + B}. \tag{5.60}$$

This representation of $r_1$ and $r_2$ is helpful because of the explicit consideration of the square root of the term $A^2/4 + B$ (this term can be real, zero or imaginary). Hence, Eq. (5.53) for $y_n$ satisfies the difference equation (5.45) provided $r_1$ and $r_2$ are given by Eqs. (5.59).

**Consistency with Initial Data.** Next, let us consider under which condition Eq. (5.53) for $y_n$ is in consistency with the initial data $y_0$ and $y_1$. To address this question we set $n = 0$ and $n = 1$ in Eq. (5.53), which leads to the conditions

$$y_0 = s + c_1 + c_2, \tag{5.61a}$$

$$y_1 = s + c_1 r_1 + c_2 r_2. \tag{5.61b}$$

A relation for $c_1$ can be found by multiplying Eq. (5.61a) with $-r_2$ and taking the sum of both equations. Similarly, a relation for $c_2$ can be found by multiplying Eq. (5.61a) with $-r_1$ and taking the sum of both equations. In this way we find the relations

$$y_1 - s - r_2(y_0 - s) = c_1(r_1 - r_2), \tag{5.62a}$$

$$y_1 - s - r_1(y_0 - s) = -c_2(r_1 - r_2). \tag{5.62a}$$

The use of these relations for $c_1$ and $c_2$ in Eq. (5.53) leads then to the solution

$$\begin{aligned} y_n &= s + \frac{y_1 - s - r_2(y_0 - s)}{r_1 - r_2} r_1^{\,n} - \frac{y_1 - s - r_1(y_0 - s)}{r_1 - r_2} r_2^{\,n} \\ &= s + \frac{r_1^{\,n} - r_2^{\,n}}{r_1 - r_2}(y_1 - s) - \frac{r_2 r_1^{\,n} - r_1 r_2^{\,n}}{r_1 - r_2}(y_0 - s) \\ &= s + \frac{r_1^{\,n} - r_2^{\,n}}{r_1 - r_2}(y_1 - s) - r_1 r_2 \frac{r_1^{\,n-1} - r_2^{\,n-1}}{r_1 - r_2}(y_0 - s). \end{aligned} \tag{5.63}$$

This solution satisfies both the difference equation (5.45) and the initial conditions that $y_n$ is equal to $y_0$ and $y_1$ at $n = 0$ and $n = 1$, respectively.

## *5.3.2 Solution Features*

**Three Solution Cases.** The solution $y_n$ will have different features depending on the value of $A^2/4 + B$ in $r_D$. We have to distinguish the cases $A^2/4 + B > 0$ ($r_D$ is real), $A^2/4 + B = 0$ ($r_D = 0$), and $A^2/4 + B < 0$ ($r_D$ is imaginary). The three cases will be discussed in the following. Regarding the first case ($A^2/4 + B > 0$), there is no problem with the direct use of Eq. (5.63) for the calculation of $y_n$. A second case is given for $A^2/4 + B = 0$. We have then $r_1 = r_2$ according to the definition (5.59) of $r_1$ and $r_2$. It is impossible to use Eq. (5.63) directly for this case because of the need to divide zero by zero. The solution can be obtained from Eq. (5.63) by taking the limit $r_D \to 0$. The latter can be done by means of l'Hospital's Rule,

$$\begin{aligned} y_n &= s + \left[\frac{d(r_1^{\,n} - r_2^{\,n})/dr_D}{d(r_1 - r_2)/dr_D}\right]_{r_D=0} (y_1 - s) - \left(\frac{A}{2}\right)^2 \left[\frac{d(r_1^{\,n-1} - r_2^{\,n-1})/dr_D}{d(r_1 - r_2)/dr_D}\right]_{r_D=0} (y_0 - s) \\ &= s + n\left[\frac{r_1^{\,n-1} + r_2^{\,n-1}}{2}\right]_{r_D=0} (y_1 - s) - (n-1)\left(\frac{A}{2}\right)^2 \left[\frac{r_1^{\,n-2} + r_2^{\,n-2}}{2}\right]_{r_D=0} (y_0 - s) \\ &= s + n\left(\frac{A}{2}\right)^{n-1} (y_1 - s) - (n-1)\left(\frac{A}{2}\right)^n (y_0 - s). \end{aligned} \tag{5.64}$$

A third case is given if $A^2/4 + B < 0$, which means that $r_D$ is an imaginary number. The eigenvalues $r_1$ and $r_2$ can be written for this case

$$r_1 = r_S + i\, r_{D^*}, \qquad r_2 = r_S - i\, r_{D^*}. \tag{5.65}$$

Here, $i$ satisfies the equation $i^2 = -1$, and we applied the abbreviation

$$r_{D^*} = \sqrt{-\frac{A^2}{4} - B}. \tag{5.66}$$

The rewriting of the solution (5.63) for this case can be handled by means of the de Moivre's formula (Abramowitz & Stegun 1984),

$$\left(r_S \pm i\, r_{D^*}\right)^n = \kappa^n \left(\cos(n\varphi) \pm i \sin(n\varphi)\right). \tag{5.67}$$

Here, $\kappa$ and $\varphi$ are defined by the expressions

$$\kappa = r_S \sqrt{1 + \left(r_{D^*} / r_S\right)^2} = \frac{r_S}{|\, r_S \,|} \sqrt{{r_S}^2 + {r_{D^*}}^2} = \frac{r_S}{|\, r_S \,|} \sqrt{-B}, \tag{5.68a}$$

$$\varphi = \arctan\left(r_{D^*} / r_S\right). \tag{5.68b}$$

The writing of $\kappa$ ensures that $\kappa = r_S$ for $r_{D^*} = 0$, which follows from Eq. (5.67) for $n = 1$. The case considered implies $B < -A^2/4$. Hence, $\kappa$ is real-valued. The use of Eq. (5.67) for ${r_1}^n$ and ${r_2}^n$ provides

$${r_1}^n - {r_2}^n = \kappa^n \left(\cos(n\varphi) + i \sin(n\varphi)\right) - \kappa^n \left(\cos(n\varphi) - i \sin(n\varphi)\right) = 2i\kappa^n \sin(n\varphi). \tag{5.69}$$

The use of the latter expression for $n$ and $n$–1 in Eq. (5.63) then implies

$$\begin{aligned} y_n &= s + \kappa^n \frac{2i \sin(n\varphi)}{2i\, r_{D^*}} (y_1 - s) - \kappa^2 \kappa^{n-1} \frac{2i \sin((n-1)\varphi)}{2i\, r_{D^*}} (y_0 - s) \\ &= s + \kappa^n \frac{\sin(n\varphi)}{r_{D^*}} (y_1 - s) - \kappa^{n+1} \frac{\sin((n-1)\varphi)}{r_{D^*}} (y_0 - s). \end{aligned} \tag{5.70}$$

In this relation we used $r_1 - r_2 = 2\, i\, r_{D^*}$ and $r_1\, r_2 = {r_S}^2 + {r_{D^*}}^2 = \kappa^2$ according to Eq. (5.65). Consequently, for the three cases considered the solution $y_n$ is given by

$$y_n = s + \begin{cases} \dfrac{{r_1}^n - {r_2}^n}{r_1 - r_2} (y_1 - s) - r_1 r_2 \dfrac{{r_1}^{n-1} - {r_2}^{n-1}}{r_1 - r_2} (y_0 - s) & \text{if } A^2/4 + B > 0 \\ n \left(\dfrac{A}{2}\right)^{n-1} (y_1 - s) - (n-1) \left(\dfrac{A}{2}\right)^n (y_0 - s) & \text{if } A^2/4 + B = 0 \\ \kappa^n \dfrac{\sin(n\varphi)}{r_{D^*}} (y_1 - s) - \kappa^{n+1} \dfrac{\sin((n-1)\varphi)}{r_{D^*}} (y_0 - s) & \text{if } A^2/4 + B < 0 \end{cases}. \tag{5.71}$$

The cases $A^2/4 + B > 0$, $A^2/4 + B = 0$, and $A^2/4 + B < 0$ will be referred to as case 1, case 2, and case 3, respectively.

**Illustration.** Characteristic features of these different solutions are illustrated in Fig. 5.4. The initial values are given by $y_0 = 0.1$ and $y_1 = 0.2$. The parameters $C$ and $B$ are given by $C = 0.1$ and $B = -B_0\, A^2/4$. The introduction of $B_0$ allows us to differentiate between the three cases considered. $A^2/4 + B$ is then given by

$$\frac{A^2}{4} + B = \left(1 - B_0\right) \frac{A^2}{4}. \tag{5.72}$$

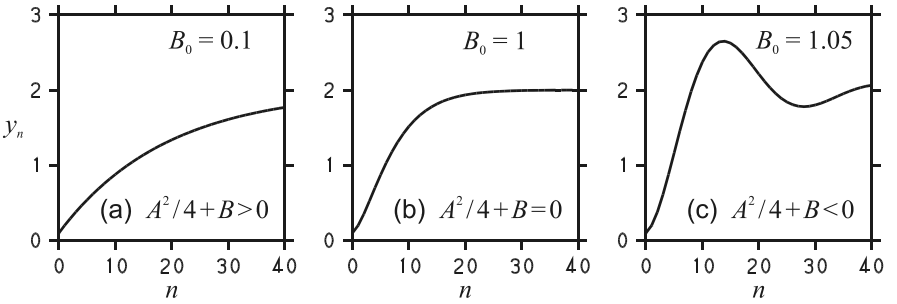



**Fig. 5.4.** An illustration of solution features of the second-order linear difference equation (5.45) for the cases 1, 2, and 3 in (a), (b), and (c), respectively.

Thus, the setting $B_0 < 1$, $B_0 = 1$, and $B_0 > 1$ corresponds to the consideration of the first, second, and third case, respectively. The parameter $A$ is chosen such that the equilibrium solution $s = 2$ is constant. This condition provides the equation

$$s = \frac{C}{1-A-B} = \frac{C}{1-A+B_0\,A^2/4} = \frac{C}{1-1/B_0+(A-2/B_0)^2\,B_0/4} \tag{5.73}$$

for the calculation of $A$. By rewriting this condition as a quadratic equation for $A$, the parameter $A$ is found to be given by

$$A = \frac{2}{B_0} - \sqrt{\frac{4}{B_0}\left(\frac{C}{s}-1+\frac{1}{B_0}\right)}. \tag{5.74}$$

The solution features are illustrated in terms of Fig. 5.4 for three values of $B_0$ that correspond to the three cases considered. Figure 5.4b shows a curve that is similar to the curve in Fig. 5.1a (which shows a first-order equation solution). A difference is given by the possibility to choose $y_1$ in the second-order equation solution. The latter fact enables the modification of the initial behavior of second-order equation solutions. Figure 5.4a shows a major difference between first-order and second-order equation solutions: the second-order equation offers the possibility to modify the transition to the equilibrium solution by the inclusion of $r_2^n$ in addition to $r_1^n$ (see the first line of (5.63)), which is often helpful for the modeling of observations. Figure 5.4c demonstrates another major difference between first-order and second-order equation solutions: the second-order equation offers the chance to involve oscillations about the equilibrium solution (which may be seen as the result of two trends that act in opposite directions). The amplitude of such oscillations depends on the value of $\kappa$: see Eq. (5.70). For $\kappa < 1$, $\kappa = 1$, and $\kappa > 1$, the oscillations are damped out, have a constant amplitude, or an increasing amplitude, respectively. The value for $\kappa$ in Fig. 5.4c is $\kappa = 0.93$. These features lead to the conclusion that the solutions of second-order equations can cover a much broader variety of cases than solutions of first-order equations.

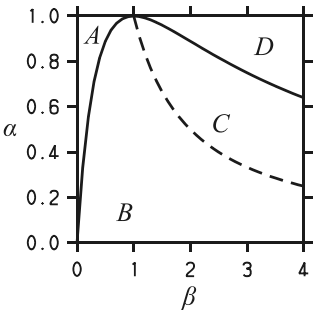



**Fig. 5.5.** Model parameter regimes that are defined by Samuelson's national income model (5.75). The functions $\alpha = 4\beta/(1+\beta)^2$ and $\alpha = 1/\beta$ are shown by the *solid* and *dashed lines*, respectively.

### *5.3.3 National Income*

**National Income Model.** Let us come back to the national income model (5.5)

$$y_n = \alpha(1+\beta)\, y_{n-1} - \alpha\beta\, y_{n-2} + 1 \tag{5.75}$$

to illustrate the advantages of using second-order difference equations. The range of variation of the marginal propensity is $0 \le \alpha \le 1$. According to Samuelson, the range of variation of $\beta$ is $0 \le \beta \le 4$, and we use $\gamma = 1$. By replacing all $y_n$ in Eq. (5.75) by $s$ and solving for $s$, we obtain for the equilibrium solution $s$

$$s = \frac{1}{1-\alpha}. \tag{5.76}$$

Hence, the equilibrium value is only controlled by α such that the transition to the equilibrium is controlled by $\beta$. Such a separation of the role of model parameters as given here is very helpful for the adjustment of a model to any given data.

**Solution.** The simplest way to find the solution for the national income model is to relate $\alpha$ and $\beta$ in (5.75) to $A$, $B$, and $C$ in Eq. (5.45). The comparison of these equations shows that

$$A = \alpha(1+\beta), \qquad B = -\alpha\beta, \qquad C = 1. \tag{5.77}$$

We calculate $A^2/4 + B$ to see which of the cases considered above is given here,

$$\frac{A^2}{4} + B = \frac{\alpha^2}{4}(1+\beta)^2 - \alpha\beta = \frac{\alpha}{4}(1+\beta)^2\left[\alpha - \frac{4\beta}{(1+\beta)^2}\right]. \tag{5.78}$$

This relation shows that all three cases considered above can be realized: $A^2/4 + B$ can be positive, negative, or equal to zero depending on the relation between $\alpha$ and $\beta$. The behavior of oscillations is determined by the value of $\kappa$. The use of the κ definition $\kappa^2 = -B$ combined with $B = -\alpha\beta$ shows that

$$\kappa = \sqrt{\alpha\beta}. \tag{5.79}$$

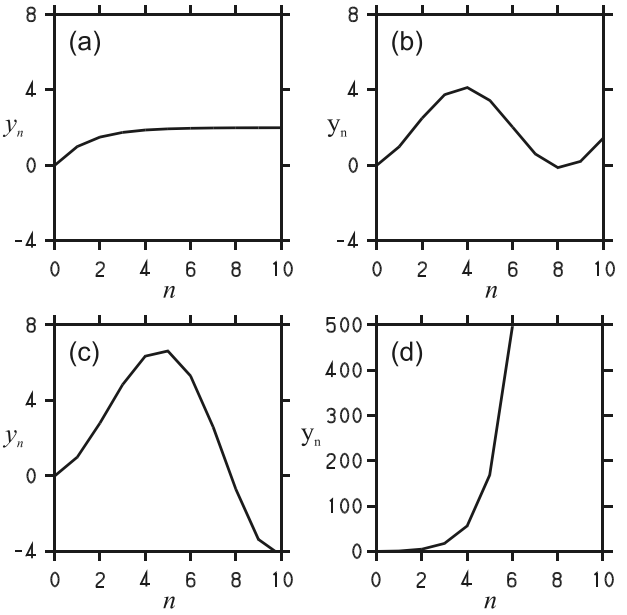



**Fig. 5.6.** Characteristic solution features of Samuelson's national income model (5.75). The model parameters applied are (a) $\alpha = 0.5$, $\beta = 0$; (b) $\alpha = 0.5$, $\beta = 2$; (c) $\alpha = 0.6$, $\beta = 2$; and (d) $\alpha = 0.8$, $\beta = 4$.

Thus, the oscillations will increase or decrease depending on whether $\alpha\beta > 1$ or $\alpha\beta < 1$, respectively. The various model parameter regimes of Samuelson's national income model (5.75) are illustrated in Fig. 5.5. The areas *A* and *D* have case-1 type solutions. The areas *B* and *C* have type-3 case solutions. The state 2 is shown by the solid line $\alpha = 4\beta/(1+\beta)^2$. The curve $\alpha = 1/\beta$ corresponds to $\kappa = 1$ that separates the regime *B* (where oscillations are damped out) from regime *C* (where the amplitude of oscillations increases).

**Solution Features.** The features of solutions of Samuelson's national income model are illustrated in Fig. 5.6, where $y_0 = 0$ and $y_1 = 1$ are applied in all cases. Samuelson identified four regimes of solution behaviors that are characterized by the corresponding $\alpha$ and $\beta$ values applied in Fig. 5.6. Figure 5.6a illustrates the case that the national income approaches asymptotically a constant level $1/(1-\alpha)$. Figure 5.6b illustrates oscillatory national income movements with a constant amplitude ($\kappa = 1$ for this case). Figure 5.6c illustrates increasing oscillatory motions ($\kappa > 1$), and Fig. 5.6d illustrates a national income with an increasing growth rate. Discussions of the economic relevance of these model features can be found elsewhere (Samuelson 1939).

### 5.3.4 Red Blood Cell Production

**First-Order Equation System.** Let us consider a model for the number of red blood cells (which carry oxygen throughout the body) circulating in the blood. The model considered is the following one (Edelstein-Keshet 2005),

$$R_n = (1-f)R_{n-1} + M_{n-1}, \tag{5.80a}$$

$$M_n = c_M\, f\, R_{n-1}. \tag{5.80b}$$

$R_n$ refers to the number of red blood cells on day $n$. Two contributions can change $R_n$. The first contribution $(1-f)\, R_{n-1}$ is proportional to the current number $R_{n-1}$ of red blood cells. The factor $1-f$ accounts for the fact that the spleen filters out and destroys a certain fraction of the red blood cells daily. Here, $f \geq 0$ is the fraction of red blood cells removed by the spleen. The second contribution $M_{n-1}$ accounts for the production of red blood cells by the bone marrow. Equation (5.80b) specifies the production $M_n$. This equation states that the production is proportional to the number of red blood cells $f\, R_{n-1}$ lost on the previous day. A parameter $c_M$ is involved in Eq. (5.80b) to account for variations of the production $M_n$.

**Second-Order Equation.** It is convenient to reformulate the model (5.80) to use the solutions of equations derived above. By taking Eq. (5.80b) at $n-1$ we find $M_{n-1} = c_M f\, R_{n-2}$. Hence, Eq. (5.80a) can be written

$$R_n = (1-f)R_{n-1} + c_M\, f\, R_{n-2}. \tag{5.81}$$

This equation is a second-order linear difference equation. It has the structure of the second-order model (5.45): for the case considered we have $y_n = R_n$, $A = 1-f$, $B = c_M f$, and $C = 0$. Hence, in difference to the national income example we have now a case where $B > 0$.

**Model Solution.** The solution to the model (5.81) can be found by means of the solution (5.71) of the difference equation (5.45), where $A = 1-f$, $B = c_M f$, and $C = 0$. According to Eqs. (5.58), the eigenvalues $r_1$ and $r_2$ are found to be given by

$$r_1 = \frac{1}{2}\left[1-f+\sqrt{(1-f)^2+4c_M\, f}\right], \quad r_2 = \frac{1}{2}\left[1-f-\sqrt{(1-f)^2+4c_M\, f}\right]. \tag{5.82}$$

For $c_M > 0$ as considered below, the term $(1-f)^2 + 4\, c_M f$ in the square root is found to be positive. Hence, we have two different and real eigenvalues $r_1$ and $r_2$. Thus, the solution to Eq. (5.81) is given by Eq. (5.63),

$$R_n = \frac{r_1^{\,n} - r_2^{\,n}}{r_1 - r_2} R_1 - r_1 r_2 \frac{r_1^{\,n-1} - r_2^{\,n-1}}{r_1 - r_2} R_0. \tag{5.83}$$

The equilibrium solution $s$ is zero here because of $C = 0$ for the case considered. The parameters $R_0$ and $R_1$ refer to $R_n$ at $n = 0$ and $n = 1$, respectively.

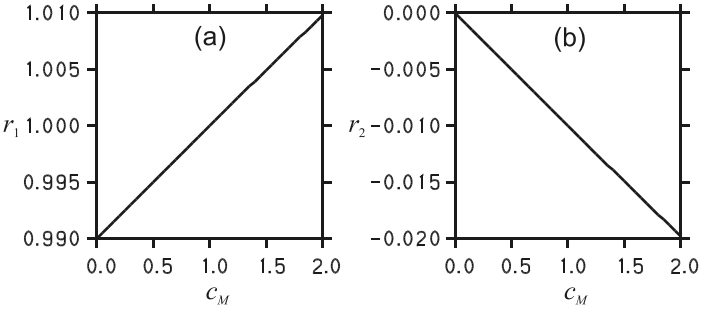



**Fig. 5.7.** The variation of the eigenvalues $r_1$ and $r_2$ with the model parameter $c_M$ according to (5.82), where $f = 0.01$ is applied.

**Production Model Specification.** The model behavior depends essentially on the eigenvalues $r_1$ and $r_2$. An example for the variation of the eigenvalues $r_1$ and $r_2$ with $c_M$ is shown in Fig. 5.7 for $f = 0.01$. It may be seen that $r_2$ is always negative and small. Hence, the effect of $r_2$ on the solution features will be rather limited: $r_2^n$ will disappear for growing $n$. The $r_1$ plot shows that $r_1$ may be smaller (for $c_M < 1$), equal (for $c_M = 1$), or larger than one (for $c_M > 1$). For $r_1$ values smaller (or larger) than one, $r_1^n$ will constantly decrease (or increase) with a growing $n$, this means the number $R_n$ of red blood cells will constantly decrease (or increase). A model that does always predict a decreasing or increasing number of red blood cells does not make a lot of sense. Thus, we will only consider the case $r_1 = 1$ (which is given for $c_M = 1$). Such a model may lead to a relatively constant number of red blood cells. We consider, therefore, in the following the model

$$R_n = (1-f)R_{n-1} + f\,R_{n-2}. \tag{5.84}$$

**Model Analysis.** For this model, Eqs. (5.82) for the eigenvalues $r_1$ and $r_2$ can be simplified in the following way,

$$r_1 = \frac{1}{2}\left[1-f+\sqrt{(1-f)^2+4f}\right] = \frac{1}{2}\left[1-f+\sqrt{(1+f)^2}\right] = 1, \tag{5.85a}$$

$$r_2 = \frac{1}{2}\left[1-f-\sqrt{(1-f)^2+4f}\right] = \frac{1}{2}\left[1-f-\sqrt{(1+f)^2}\right] = -f. \tag{5.85b}$$

According to Eq. (5.83), the solution for $R_n$ is given for this case by

$$\begin{aligned} R_n &= \frac{1-(-f)^n}{1+f}R_1 + f\,\frac{1-(-f)^{n-1}}{1+f}R_0 = \frac{R_1}{1+f} + \frac{f\,R_0}{1+f} + \frac{(-f)^n}{1+f}\left(R_0-R_1\right) \\ &= \frac{R_1}{1+f} + \frac{(1+f-1)\,R_0}{1+f} + \frac{(-f)^n}{1+f}\left(R_0-R_1\right) = R_0 + \frac{R_1-R_0}{1+f} - (-f)^n\,\frac{R_1-R_0}{1+f}. \end{aligned} \tag{5.86}$$

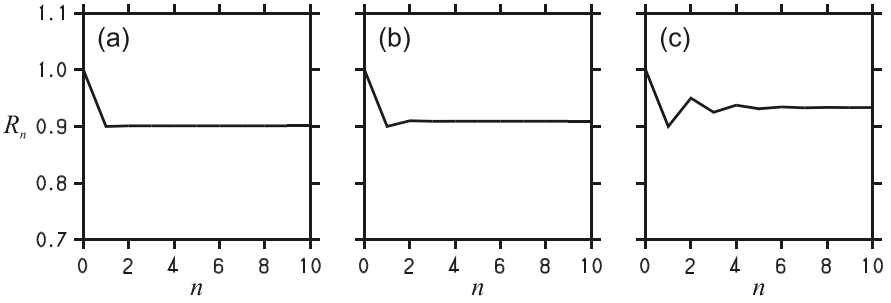



**Fig. 5.8.** Solutions $R_n$ of the red blood cell production model (5.84), where the initial values $R_0 = 1$ and $R_1 = 0.9$ are used. In (a), (b), and (c), the values $f = 0.01$, $f = 0.1$, and f = 0.5 are used, respectively.

For large $n$, $(-f)^n$ will disappear. Hence, the equilibrium solution $R_e$ reads

$$R_e = R_0 + \frac{R_1 - R_0}{1 + f}. \tag{5.87}$$

It is worth noting that this expression cannot be found immediately by replacing $R_n$ by $R_e$ in Eq. (5.84): we end up with the identity $R_e = R_e$ in this way. Regarding the following discussion it is helpful to rewrite the solution (5.86) of the red blood cell model (5.84) by taking reference to the equilibrium solution (5.87),

$$R_n = R_e - (-f)^n \left(R_e - R_0\right). \tag{5.88}$$

Due to the appearance of the factor $(-f)^n$, the solution $R_n$ does oscillate about the equilibrium solution $R_e$ until $R_e$ is reached. The reason for these oscillations is the time delay considered: the bone marrow produces red blood cells proportional to the number of cells lost on the previous day.

**Model Applicability.** The suitability of the model (5.84) is considered in terms of Fig. 5.8. For a reasonable value $f = 0.01$, this figure shows that the transition to the equilibrium value takes place very fast (see Fig. 5.8a). At $n = 1$, there is only a very little difference to the equilibrium solution ($R_1 = 0.9$ and $R_e = 0.901$). At $n = 2$, we have, basically, the equilibrium value ($R_2 = R_e + 9.9 \times 10^{-6}$). Such a model behavior is clearly unreasonable. It requires unreasonable values of $f = 0.1$ and $f = 0.5$ to slow down the fast transition to the equilibrium. Another interesting finding is that the model (5.84) cannot respond to perturbations. Let us consider the value $R_1 = 0.9$ as a perturbation (which may be the result of a donation of blood). Then, there is no homeostasis mechanism such that $R_n$ can return to the previous value of one. Due to these reasons, the model (5.84) represents a poor model for the production of red blood cells. It is worth noting that is requires much more complex models to deal with reality (Yafia 2009).

## 5.4 Nonlinear Changes

The third equation type that we will discuss is a nonlinear first-order difference equation. The initial value $y_0$ at $n = 0$ is considered to be given. The $y_n$ values for $n = 1, 2, \ldots$ are determined by the equation

$$y_n = f(y_{n-1}). \tag{5.89}$$

Here, the right-hand side $f(y_{n-1})$ represents any given (linear or nonlinear) function of $y_{n-1}$. The addition of $- y_{n-1}$ on both sides shows that this equation determines changes $y_n - y_{n-1}$. Equation (5.89) is nonhomogeneous if $f(y_{n-1})$ adds any constant (as given, for example, for the case $f(y_{n-1}) = y_{n-1}^2 + 1$). We have a first-order difference equation because only function values at $n$ and $n - 1$ are involved.

### *5.4.1 Analysis Concepts*

A significant difference between linear and nonlinear difference equations is that analytical solutions of nonlinear equations are usually unavailable. Hence, we have to use other methods to analyze such equations.

**Pendulum.** Let us consider an example in order to illustrate the basic concepts. Figure 5.9 illustrates the motion of a pendulum, where α denotes the angle that the pendulum makes with the vertical direction. A detailed mathematical explanation of the pendulum motion is not really simple: it requires the use of a nonlinear differential equation (see the discussion of this problem in Sect. 9.4). Nevertheless, we can determine the characteristic features of pendulum motions without the application of relatively advanced mathematical tools. First of all, it is helpful to see which final state the system will realize. The latter question can be addressed in two steps. The first step is to find equilibrium solutions (this means stationary or steady state solutions). Figure 5.9 shows that there are two equilibrium positions of the pendulum motion: the position 1 with the mass directly below the point of support (i.e., for $\alpha = 0$), and the position 2 with the mass directly above the point of support (i.e., for $\alpha = \pi$). The second step is to clarify whether the system will realize the potential equilibrium solutions.

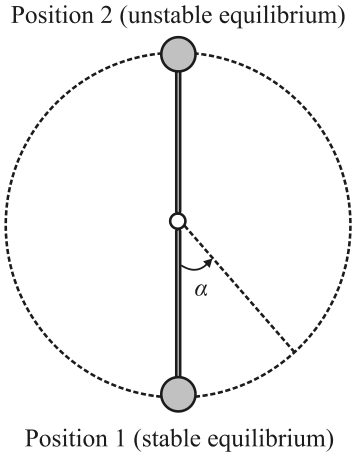



**Fig. 5.9.** A pendulum.

The best way to find the answer to this question is to consider motions relatively close to the equilibrium points and to ask whether the equilibrium points will be realized or not. An equilibrium point is called stable if the system will realize this equilibrium point, otherwise the equilibrium point is called unstable. With regard to the equilibrium points shown in Fig. 5.9 we see a different behavior. If the mass is slightly displaced from the lower equilibrium position 1, the pendulum will oscillate back and forth with gradually decreasing amplitude until the equilibrium point is finally realized. Such an equilibrium point is stable (it attracts the motion). If the mass is slightly displaced from the upper equilibrium point 2, it will rapidly fall and the mass will finally approach the equilibrium point 1. Thus, the second equilibrium point is unstable.

**Equilibrium.** How is it possible to find equilibrium solutions of any system? A steady state means that $y_n$ does not depend on $n$ anymore, this means we have $y_n = y_{n-1} = y_e$, where $y_e$ refers to the equilibrium solution. According to $y_n = f(y_{n-1})$, the condition for finding $y_e$ is given by the equation

$$y_e = f(y_e). \tag{5.90}$$

This equation may well have several solutions if we have to deal with a nonlinear equation.

**Stability.** How is it possible to decide whether any equilibrium point is stable or not? To address this question we have to consider $y_n$ relatively close to any equilibrium point,

$$y_n = y_e + u_n. \tag{5.91}$$

Here, $u_n$ represents a little deviation from the equilibrium solution. By making use of this expression in $y_n = f(y_{n-1})$ we find

$$y_e + u_n = f(y_e + u_{n-1}). \tag{5.92}$$

Now, we account for the fact that $u_n$ is small. In particular, we expand the right-hand side in terms of the Taylor series about $y_e$, where terms that are nonlinear in $u_n$ are neglected,

$$y_e + u_n = f(y_e) + f'(y_e)u_{n-1}. \tag{5.93}$$

Due to the definition $y_e = f(y_e)$ of the equilibrium solution, this equation reduces to

$$u_n = f'(y_e)u_{n-1}. \tag{5.94}$$

This equation is a homogeneous linear first-order difference equation. According to the discussion in Sect. 5.2, $u_n$ will disappear for large $n$ if

$$\left|f'(y_e)\right| < 1. \tag{5.95}$$

This inequality represents the condition for a stable equilibrium point: the solution will approach the equilibrium point if this inequality is satisfied.

### *5.4.2 The Discrete Logistic Equation: Analysis*

**Discrete Logistic Equation.** As an example, let us consider a quadratic first-order difference equation

$$y_n = A\,y_{n-1} + B\,y_{n-1}{}^2 + C, \tag{5.96}$$

where $A$, $B$, and $C$ are any model parameters. In particular, we will assume that $A = 1 + a$, $B = -a/K$, and $C = 0$,

$$y_n = (1+a)\,y_{n-1} - \frac{a}{K}\,y_{n-1}{}^2 = y_{n-1} + a\,y_{n-1}\left(1 - \frac{y_{n-1}}{K}\right). \tag{5.97}$$

Equation (5.97) is called the discrete logistic equation or the logistic difference equation. The assumptions for $A$ and $B$ represent appropriate rewritings of $A$ and $B$ in terms of two other parameters $a$ and $K$, and the assumption $C = 0$ simplifies the following analysis. A closer look at Eq. (5.97) reveals the suitability of the assumptions $A = 1 + a$ and $B = -a/K$. We see that one equilibrium solution is given for the case that $y_{n-1} = K$: the parenthesis term is zero for this case, which means that $y_n = y_{n-1}$ for all n. A second equilibrium solution is given for $y_{n-1} = 0$, which does also result in $y_n = y_{n-1}$ for all $n$. We are interested to ensure that $y_n \geq 0$ for all $n$ such that the model considered can be used, for example, as a population model. Thus, we assume that $y_0 \geq 0$, and we also assume that $a > 0$ and $K > 0$. The consideration of negative $K$ values would mean that $y_n$ can realize a negative equilibrium solution. The consideration of negative $a$ values would mean that $y_n$ will approach zero or infinity for large $n$ depending on whether the parenthesis term in Eq. (5.97) is positive or negative initially, respectively (a negative $a$ value can also imply oscillations of $y_n$ about zero including the appearance of negative $y_n$ values). Thus, the equilibrium value $K$ cannot be realized if $a < 0$.

**Nondimensional Discrete Logistic Equation.** Instead of working directly with Eq. (5.97), it is helpful to analyze a nondimensional version of this model. For doing this, we write the middle expression in Eq. (5.97) as

$$y_n = (1+a)\,y_{n-1}\left[1 - \frac{a}{(1+a)K}\,y_{n-1}\right]. \tag{5.98}$$

Now, we introduce the variable

$$x_n = \frac{a}{(1+a)K}\,y_n. \tag{5.99}$$

The variable $x_n$ is nondimensional because it is compared to one in the bracket term of Eq. (5.98). By multiplying Eq. (5.98) with $a/[(1+a)\,K]$ we obtain

$$x_n = f(x_{n-1}) = (1+a)\,x_{n-1}(1 - x_{n-1}). \tag{5.100}$$

This equation is called the nondimensional discrete logistic equation.

**Parameter Variations.** We are interested to ensure that $x_n \geq 0$ for all $n$, which implies that $y_n \geq 0$ for all $n$. The condition $x_n \geq 0$ requires that $x_{n-1} \leq 1$: otherwise $x_n$ will become negative according to Eq. (5.100). To see the upper bound of $x_{n-1}$, we need to know the maximum of $x_n$ for all $n$. This maximum can be obtained by calculating the maximum of $f(x_{n-1})$. The derivatives of $f(x_{n-1})$ are given by

$$\frac{df}{dx_{n-1}} = (1+a)(1-2x_{n-1}), \qquad \frac{d^2 f}{dx_{n-1}^{\ 2}} = -2(1+a). \tag{5.101}$$

Hence, $f(x_{n-1})$ has a maximum value at $x_{n-1} = 1/2$. The maximum of $f(x_{n-1})$ at $x_{n-1} = 1/2$ is given by

$$f(1/2) = \max x_n = (1+a)/4. \tag{5.102}$$

Consequently, the condition $a \leq 3$ ensures that max $x_n \leq 1$ for all $n$. As a consequence of this analysis, the variation of $a$ to be considered is given by

$$0 < a \leq 3. \tag{5.103}$$

This condition ensures $0 \leq x_n \leq 1$ for all $n$.

**Equilibrium.** To find the equilibrium solutions of the nondimensional discrete logistic equation, we replace $x_n$ and $x_{n-1}$ by $x_e$ in (5.100),

$$x_e = (1+a)x_e(1-x_e) = x_e + a x_e - (1+a)x_e^{\ 2}. \tag{5.104}$$

The corresponding quadratic equation for $x_e$ can be written

$$0 = x_e\big(a - (1+a)x_e\big). \tag{5.105}$$

Hence, we find two equilibrium solutions $x_e^{(1)}$ and $x_e^{(2)}$,

$$x_e^{\ (1)} = 0, \qquad x_e^{\ (2)} = \frac{a}{1+a}. \tag{5.106}$$

**Stability.** The stability of these equilibrium points can be evaluated according to the inequality (5.95). According to Eq. (5.101), the derivative $f'$ at the equilibrium points $x_e^{(1)}$ and $x_e^{(2)}$ is given by

$$f'(x_e^{\ (1)}) = 1+a, \qquad f'(x_e^{\ (2)}) = (1+a)\left(1 - 2\frac{a}{1+a}\right) = 1 + a - 2a = 1 - a. \tag{5.107}$$

The stability condition $|f'(x_e)| < 1$ then implies for the equilibrium solutions $x_e^{(1)}$ and $x_e^{(2)}$ the stability conditions

$$x_e^{\ (1)}: \quad -2 < a < 0, \qquad x_e^{\ (2)}: \quad 0 < a < 2. \tag{5.108}$$

The first equilibrium solution $x_e^{(1)}$ is an unstable equilibrium point for the $a$ values $0 < a \leq 3$ considered. Regarding $x_e^{(2)}$ we find the following behavior,

$$0 < a < 2: \quad \text{stable} \qquad 2 \leq a \leq 3: \quad \text{unstable} \tag{5.109}$$

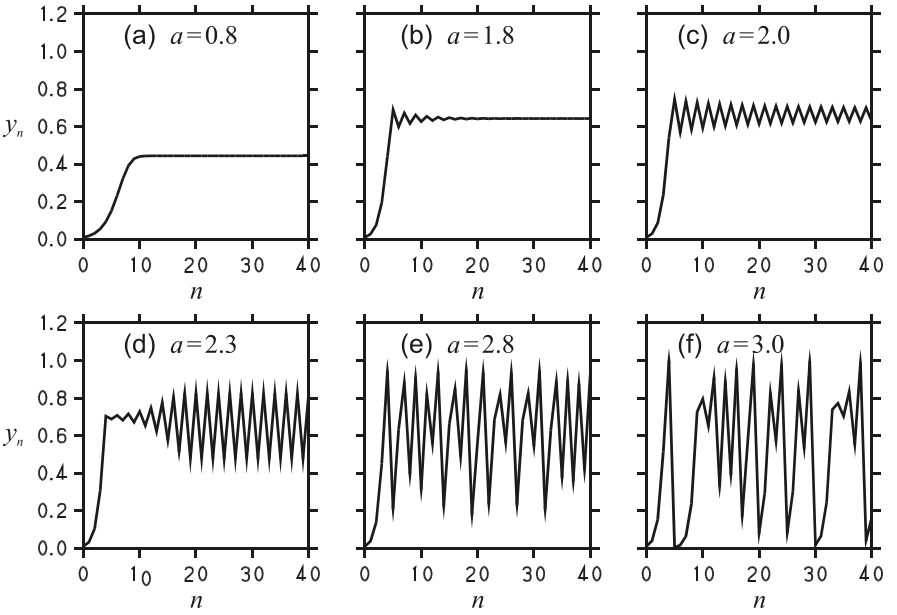



**Fig. 5.10.** An illustration of solutions of the nondimensional discrete logistic equation (5.100) in dependence on $n$, where $y_0 = 0.01$. The corresponding values of the model parameter $a$ are shown in the figures.

At $a = 2$ the stability properties of the equilibrium solution $x_e^{(2)}$ change. The values of $a$ for which such changes occur are called bifurcation values.

### *5.4.3 The Discrete Logistic Equation: Solutions*

The features of numerical solutions of the nondimensional discrete logistic equation (5.100) are illustrated in Fig. 5.10, where $y_0 = 0.01$ is applied. There are several solution regimes.

**Stable Growth.** One solution regime is given for $0 < a < 2$. For relatively low $a$ values, $0 < a < 1$, we observe a smooth solution. Initially, there is an exponential increase, but then the curve levels off. The asymptotic value is determined by $x_e^{(2)} = a/(1 + a)$. For larger values $1 < a < 2$ the increase is so rapid that the solution overshoots the equilibrium value. A period of damped oscillations follows before the solution converges to the equilibrium value $x_e^{(2)} = a/(1 + a)$. The solution behavior for the value $a = 2$ is shown in Fig. 5.10c. This figure shows again damped oscillations, but the equilibrium point $x_e^{(2)} = 2/3$ cannot be realized for finite $n$ (which means that the solution shows an unstable behavior).

**Cyclic Growth.** Another solution behavior is observed for $2 < a < 2.57$. The solution shows oscillations which are no longer damped: after an initial stage the solution shows bounded oscillations. In Fig. 5.10d, the bounds are given by 0.4794 and 0.8236. The solution has a cycle: after the initial stage the solution repeats itself after two intervals (i.e., the solution is the same between $n = 30$ and $n = 32$, and between $n = 32$ and $n = 34$). This cycle is called a 2-cycle. For $a = 2.5$, the cycle changes to a 4-cycle (for every four values the solution is the same), and there are higher-order cycles for higher values of $a$.

**Chaotic Growth.** However, there is a limit for such cyclic growth at $a \approx 2.57$: for higher values of $a$ there is no cyclic growth, but the growth patterns become random, i.e., we find a chaotic solution behavior for $2.57 < a \leq 3$. Figure 5.10f illustrates the problem with values of $a$ bigger than 3: the solution may become negative in these cases. The first systematic studies about the chaotic behavior of solutions of the logistic difference equation where made by May (1975, 1976). The number $a \approx 2.57$ was discovered by Feigenbaum (1978). Thus, this value is called the Feigenbaum number. A nice discussion of investigations of properties of the logistic difference equation can be found in the book of Gleick (1987). The textbooks of Edelstein-Keshet (2005) and Allen (2007) also provide very helpful insight.

## 5.5 Difference and Differential Equations

The purpose of this section is to illustrate some advantages and disadvantages of difference equations. It will be shown that the use of difference equations faces several problems related, for example, to the parameters of difference equations, the interpolation between data values, and the analysis of equations and solutions. Such problems can be overcome by the application of differential equations. The latter does not mean, however, that difference equations are not useful. Instead, it will be shown that difference equations play a very important role: they are used for the formulation of differential equations, the numerical solution of differential equations, and in some cases even the analytical solution of differential equations. The common use of differential and difference equations is such that laws (e.g., the laws of mechanics, population ecology, and probability theory: see the discussion of these equations in Chaps. 7–10) are usually formulated in terms of differential equations. Such laws are often represented in terms of nonlinear equations that cannot be solved analytically. The application of difference equations is then the only way to find numerical solutions to such nonlinear differential equations. Therefore, difference equations definitely represent a very valuable concept.

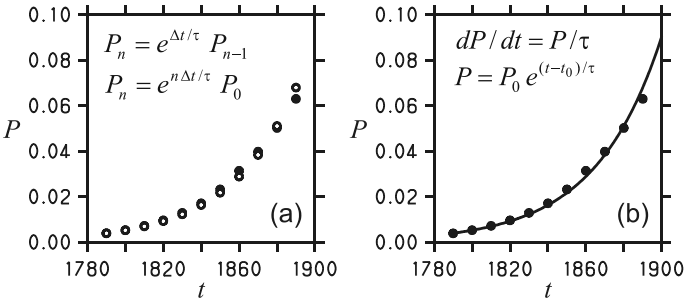



**Fig. 5.11.** A comparison between difference and differential equations regarding the U.S. population development between 1790 and 1890. The population data are shown as *solid symbols*. The *open symbols* show the solution of a difference equation, which depends on the distance $\Delta t$ of data points. The *solid line* shows the solution of a differential equation, which does only depend on the process parameter $\tau$ and initial data.

### *5.5.1 Difference Versus Differential Equations*

**U.S. Population Data.** Let us consider again the U.S. population data from 1790 to 1890 discussed in Chap. 1 to illustrate a few questions regarding the use of difference equation. The observed population data are presented in Fig. 5.11. According to expression (1.19), the U.S. population data can be well described by the function

$$P = P_0\, e^{\frac{t-t_0}{\tau}}. \tag{5.110}$$

The parameters of this exponential model are given by $P_0 = 0.0039$, $t_0 = 1790$, and the characteristic time scale is $\tau = 35$ years.

**Difference Equation Model.** First, let us design a difference equation model in agreement with the U.S. population data. A simple way to derive this model is to rewrite the known model (5.110). We introduce $t - t_0 = n\,\Delta t$. Here, $\Delta t$ is the time step considered, which is 10 years for the U.S. population data, and $n = 0, 1, \ldots$. The exponential population model can be written then

$$P_n = P_0\, a^n, \tag{5.111}$$

where $a = \exp(\Delta t / \tau)$. This expression represents the solution of a first-order linear difference equation that is given by (see Sect. 5.2)

$$P_n = a\, P_{n-1}. \tag{5.112}$$

**Differential Equation Model.** A differential equation model can be found correspondingly. The differentiation of (5.110) leads to

$$\frac{dP}{dt} = \frac{P_0}{\tau} e^{\frac{t-t_0}{\tau}}. \tag{5.113}$$

The use of $P$ given by (5.110) on the right-hand side enables the rewriting

$$\frac{dP}{dt} = \frac{P}{\tau}. \tag{5.114}$$

This equation represents a differential equation model for $P$. The solution for this model is given by the exponential model (5.110).

**Problems of Difference Equations.** The comparison of the solution (5.111) of the difference model (5.112) with the U.S. population data in Fig. 5.11a reveals several problems. First, in reality we often have continuous processes, but observations about these processes are only given for a certain resolution (i.e., for a certain $\Delta t$ that is ten years for the population data considered). Thus, usually we are interested to have information about the process between the observed data points. A difference equation that is designed such that it agrees with observed data does not provide such insight. Second, it is often helpful to be able to use models analytically to calculate, e.g., the minimum or maximum of functions. The structure of the solution of difference models makes it usually difficult to use such solutions analytically. Third, actually we look for models that are as universal as possible, this means we are interested in models that can be applied with the same model parameters to many different cases. Difference equations do not represent such universal equations: we have to use different equations for every distance between data values. The latter can be seen by considering the expressions (5.111) and (5.112), which depend on $\Delta t$. By adjusting the difference equation to the data set as considered in Chap. 1 (data values at 1790, 1800, 1810, …), we have $a = \exp(10/35) \approx 1.33$ for $\Delta t = 10$ years and $\tau = 35$ years. Alternatively, we could use the exponential model (5.110) such that the distance of function values is only one year (values at 1790, 1791, 1792, …). For this case, we have $\Delta t = 1$ such that the parameter $a$ would be $a = \exp(1/35) = 1.029$. In general, we could try to fix this problem by accounting for the resolution effect (the setting of $\Delta t$) on the model parameters. Such an approach may work for a simple case as considered here, but it is hard to use for more complex cases.

**Difference Versus Differential Equations.** These problems can be overcome by the use of differential equations: see the illustration in Fig. 5.11b. Differential equations provide continuous solutions for all t, it is often possible to use such solutions for analytical investigations, and the parameters of differential equations are independent of the resolution of processes. Nevertheless, this does not mean that difference equations do not represent a valuable concept. Difference equations

are used for the formulation of differential equations, their numerical solution, and even their analytical solution. These advantages of difference equations will be illustrated in the next two subsections regarding the solution of first- and second-order linear differential equations.

### *5.5.2 First-Order Difference and Differential Equations*

**Problem.** Let us consider first a linear first-order differential equation

$$\frac{dy}{dt} = -\frac{y - y_e}{T}. \tag{5.115}$$

The constant $y_e$ represents the equilibrium value of $y$ (for $y = y_e$ there is no change of $y$ anymore). The constant $T$ refers to a characteristic time scale that determines the transition between any initial value $y(0)$ and $y_e$. A discussion of the physical relevance of this equation can be found in Sect. 7.2. Equation (5.115) is defined as limit $\Delta t \to 0$ of the difference equation

$$\frac{y_n - y_{n-1}}{\Delta t} = -\frac{y_{n-1} - y_e}{T}. \tag{5.116}$$

This equation represents the first-order linear difference equation (5.6), where $a = -\Delta t / T$ and $b = y_e \, \Delta t / T$. In the following, we will determine the analytical solution of Eq. (5.115) on the basis of the difference equation (5.116), and we will show that the numerical solution of the differential equation (5.115) can be found by means of the difference equation (5.116) if a sufficiently small $\Delta t$ is used.

**Analytical Solution.** The solution (5.15) of the difference equation (5.6) shows that the solution of the Eq. (5.116) is given by

$$y_n = \frac{b}{1-a} + a^n \left( y_0 - \frac{b}{1-a} \right) = y_e + \left( 1 - \frac{\Delta t}{T} \right)^n (y_0 - y_e). \tag{5.117}$$

The last expression makes use of $a = 1 - \Delta t / T$ and $b = y_e \, \Delta t / T$, and $y_0$ refers to the initial value of $y_n$ (as may be seen by setting $n = 0$). To find the limit $\Delta t \to 0$ of this expression we consider first

$$\lim_{\Delta t \to 0} \left( 1 - \frac{\Delta t}{T} \right)^n = \lim_{\Delta t \to 0} \exp\left\{ n \ln\left( 1 - \frac{\Delta t}{T} \right) \right\} = \lim_{\Delta t \to 0} \exp\left\{ -\frac{n \, \Delta t}{T} \right\} = e^{-t/T}. \tag{5.118}$$

Here, $t = n \, \Delta t$ is used, and the logarithmic function is approximated by its Taylor series in the first order of approximation,

$$\ln\left( 1 - \frac{\Delta t}{T} \right) = -\frac{\Delta t}{T}. \tag{5.119}$$

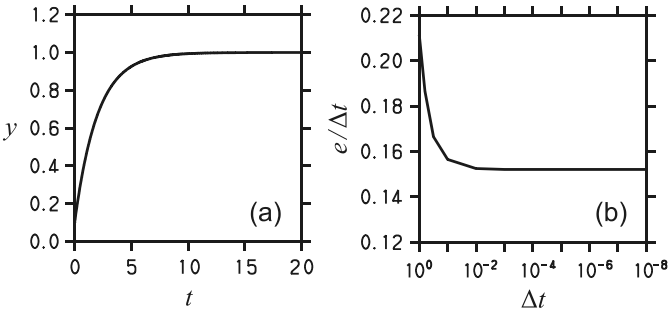



**Fig. 5.12.** The convergence of numerical solutions of the differential equation (5.121); (a) the analytical solution; (b) the normalized error $e / \Delta t$ of numerical solutions of Eq. (5.121) in dependence on the time step $\Delta t$.

Hence, the solution (5.117) reads in the limit $\Delta t \to 0$

$$y(t) = y_e + e^{-t/T}(y_0 - y_e), \tag{5.120}$$

where $y_n$ is replaced by the continuous function $y(t)$.

**Numerical Solution.** It is often the case that analytical solutions of differential equations cannot be found. In such a case, we have to find numerical solutions of differential equations via the solution of a difference equation. To illustrate the correctness of such numerical solutions for a sufficiently small $\Delta t$, we consider the equation

$$\frac{dy}{dt} = -\frac{y-1}{2} \tag{5.121}$$

combined with the initial condition $y_0 = 0.1$. The analytical solution (5.120) of this equation is shown in Fig. 5.12a. The numerical solution of Eq. (5.121), which is obtained via the difference equation (5.116), depends on the time step $\Delta t$. To assess the error of the numerical solution we introduce the standardized error

$$e = \max\left|\frac{y_n - y(t)}{y(t)}\right|. \tag{5.122}$$

This error considers the absolute value of the normalized difference between the solution $y_n$ of the difference equation (5.116) and the analytical solution $y(t)$ given by (5.120). The maximum is taken over of all absolute values with $0 \le t \le 20$. Fig. 5.12b shows that the normalized error $e / \Delta t$ is found to be independent of $\Delta t$ for increasing $\Delta t$. In particular, we find

$$e = 0.1521\,\Delta t \tag{5.123}$$

for $\Delta t \le 10^{-3}$. Hence, the solution of the difference equation (5.116) represents an accurate tool to solve the differential equation (5.115) numerically.

### 5.5.3 Second-Order Difference and Differential Equations

**Problem.** Next, let us consider the linear second-order differential equation

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + c\,y = 0, \tag{5.124}$$

where $a$, $b$, and $c$ are any model constants. This equation will be used in Chap. 7 for the explanation of laws in mechanics. We define this equation as the limit $\Delta t \to 0$ of the difference equation

$$0 = \frac{a}{\Delta t}\left\{\frac{y_n - y_{n-1}}{\Delta t} - \frac{y_{n-1} - y_{n-2}}{\Delta t}\right\} + b\,\frac{y_{n-1} - y_{n-2}}{\Delta t} + c\,y_{n-2}. \tag{5.125}$$

The first term represents the second-order derivative defined as derivative of first-order derivatives. The second term can be justified by setting $c = 0$. For this case, Eq. (5.125) provides a correct first-order equation for $x_n = (y_n - y_{n-1})/\Delta t$. The third term can be justified by setting $a = 0$, which results in a correct first-order equation. To compare Eq. (5.125) with the second-order difference equation (5.45) we write Eq. (5.125) as

$$0 = y_n - 2y_{n-1} + y_{n-2} + \frac{b}{a}\Delta t\,(y_{n-1} - y_{n-2}) + \frac{c}{a}\Delta t^2\,y_{n-2}, \tag{5.126}$$

where $a \neq 0$ is assumed. The latter equation can be written

$$y_n = A\,y_{n-1} + B\,y_{n-2}, \tag{5.127}$$

where $A$ and $B$ are given by

$$A = 2 - \frac{b}{a}\Delta t, \qquad B = -1 + \frac{b}{a}\Delta t - \frac{c}{a}\Delta t^2. \tag{5.128}$$

**Reformulated Problem.** The second-order difference equation (5.127) can be also formulated in terms of a system of two first-order difference equations, which will be relevant regarding the discussion below. To obtain this system of two first-order difference equations we consider the variable $x_n$ that is defined by

$$\frac{y_n - y_{n-1}}{\Delta t} = x_{n-1}. \tag{5.129}$$

By using this definition of $x_n$ we can write Eq. (5.125) as

$$0 = a\,\frac{x_{n-1} - x_{n-2}}{\Delta t} + b\,x_{n-2} + c\,y_{n-2}. \tag{5.130}$$

We can replace $n - 1$ by $n$ in this equation. The combination of the corresponding equation with Eq. (5.129) results in a system of first-order difference equations,

which is given by

$$\frac{y_n - y_{n-1}}{\Delta t} = x_{n-1}, \tag{5.131a}$$

$$\frac{x_n - x_{n-1}}{\Delta t} = -\frac{c}{a} y_{n-1} - \frac{b}{a} x_{n-1}. \tag{5.131b}$$

**Analytical Solution.** To find the analytical solution of the differential equation (5.124) on the basis of the solution of the difference equation (5.127) we have to calculate the eigenvalues $r_1$ and $r_2$ related to Eq. (5.127). Therefore we consider

$$\frac{A^2}{4} + B = \frac{1}{4}\left[4 - 4\frac{b}{a}\Delta t + \frac{b^2}{a^2}\Delta t^2\right] - 1 + \frac{b}{a}\Delta t - \frac{c}{a}\Delta t^2 = \frac{\Delta t^2}{4a^2}(b^2 - 4ac). \tag{5.132}$$

According to Eq. (5.58), the eigenvalues $r_1$ and $r_2$ are then given by

$$\begin{aligned} r_1 &= 1 - \frac{b}{2a}\Delta t + \frac{\Delta t}{2}\sqrt{\frac{b^2 - 4ac}{a^2}} = 1 - \frac{b}{2a}\Delta t \pm \frac{\Delta t}{2a}\sqrt{b^2 - 4ac}, \\ r_2 &= 1 - \frac{b}{2a}\Delta t - \frac{\Delta t}{2}\sqrt{\frac{b^2 - 4ac}{a^2}} = 1 - \frac{b}{2a}\Delta t \mp \frac{\Delta t}{2a}\sqrt{b^2 - 4ac}. \end{aligned} \tag{5.133}$$

Here, the positive sign in the $r_1$ expression (the negative sign in the $r_2$ expression) applies to the case that the parameter $a$ is positive. Consequently, we obtain two eigenvalues: one eigenvalue (which we will call $r_1$) involves the square root with a positive sign, and another eigenvalue (which we will call $r_2$) involves the square root with a negative sign. Hence, the eigenvalues read

$$r_1 = 1 + s_1 \Delta t, \qquad r_2 = 1 + s_2 \Delta t, \tag{5.134}$$

where $s_1$ and $s_2$ are defined by

$$s_1 = \frac{1}{2a}\left[-b + \sqrt{b^2 - 4ac}\right], \qquad s_2 = \frac{1}{2a}\left[-b - \sqrt{b^2 - 4ac}\right]. \tag{5.135}$$

According to Eq. (5.63) the solution of the equation (5.127) is then given by

$$y_n = \frac{y_1 - r_2 y_0}{r_1 - r_2} r_1^{\,n} - \frac{y_1 - r_1 y_0}{r_1 - r_2} r_2^{\,n}. \tag{5.136}$$

By using the relations $r_1 = 1 + s_1 \Delta t$ and $r_2 = 1 + s_2 \Delta t$ we can write the coefficients of $r_1^{\,n}$ and $r_2^{\,n}$ in this solution as

$$\begin{aligned} \frac{y_1 - (1 + s_2 \Delta t) y_0}{(s_1 - s_2)\Delta t} &= \frac{y_1 - y_0 - s_2 \Delta t\, y_0}{(s_1 - s_2)\Delta t} = \frac{y'_0\, \Delta t - s_2 \Delta t\, y_0}{(s_1 - s_2)\Delta t} = \frac{y'_0 - s_2 y_0}{s_1 - s_2}, \\ \frac{y_1 - (1 + s_1 \Delta t) y_0}{(s_1 - s_2)\Delta t} &= \frac{y_1 - y_0 - s_1 \Delta t\, y_0}{(s_1 - s_2)\Delta t} = \frac{y'_0\, \Delta t - s_1 \Delta t\, y_0}{(s_1 - s_2)\Delta t} = \frac{y'_0 - s_1 y_0}{s_1 - s_2}, \end{aligned} \tag{5.137}$$

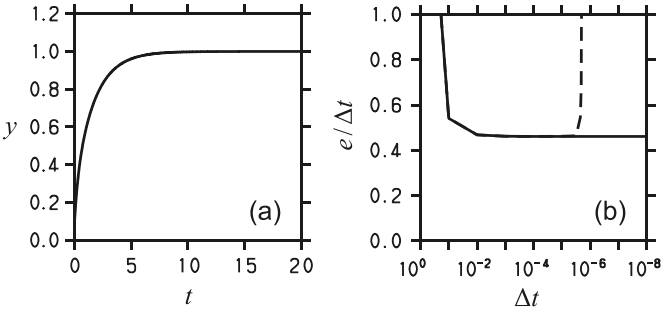



**Fig. 5.13.** The convergence of numerical solutions of the differential equation (5.142); (a) the analytical solution; (b) the *dashed line* shows the normalized error $e / \Delta t$ of numerical solutions of the discrete second-order equation (5.127); the *solid line* in (b) shows the error of numerical solutions of the system (5.131) of two discrete first-order equations.

where the abbreviation $y'_0 = (y_1 - y_0) / \Delta t$ is introduced for the initial derivative. Hence, the solution of the second-order difference equation (5.127) reads

$$y_n = \frac{y'_0 - s_2 y_0}{s_1 - s_2} r_1^n - \frac{y'_0 - s_1 y_0}{s_1 - s_2} r_2^n . \tag{5.138}$$

In the limit $\Delta t \to 0$, we can use the first-order approximation of the Taylor series of exponential functions to replace $r_1 = 1 + s_1 \Delta t$ and $r_2 = 1 + s_2 \Delta t$ by

$$r_1 = e^{s_1 \Delta t}, \qquad r_2 = e^{s_2 \Delta t}. \tag{5.139}$$

Therefore, we find the solution (5.138) in the limit $\Delta t \to 0$ to be given by

$$y = \frac{y'_0 - s_2 y_0}{s_1 - s_2} e^{s_1 t} - \frac{y'_0 - s_1 y_0}{s_1 - s_2} e^{s_2 t}. \tag{5.140}$$

Here, $t = n \Delta t$ is used, and $y_n$ is replaced by the continuous function $y(t)$. This solution is real if $b^2 - 4\, a\, c > 0$ such that $s_1$ and $s_2$ are real. In correspondence to the solutions (5.71) of the second-order difference equation (5.45) it is possible to obtain also real solutions for $b^2 - 4\, a\, c \le 0$ by a rewriting of Eq. (5.140): see Sect. 5.3.2. Regarding the following discussion it is interesting to note that the solution (5.140) does only allow zero or infinite asymptotic solutions depending on whether $s_2$ and $s_2$ are negative or positive. Equations (5.135) show that the case $s_1 = 0$ and $s_2 < 0$, which would provide a solution that levels off at a nonzero value, is only given for $a = 0$ or $c = 0$, i.e., we would not consider a second-order differential equation in this case.

**Numerical Solution.** As done with regard to the linear first-order differential equation considered before, let us illustrate how a linear second-order differential equation can be solved numerically by means of difference equations. To have a fair comparison to the result obtained above for the numerical solution of a first-order differential equation we are interested in a solution that looks similar to the solution of the first-order differential equation (a solution that levels off at one). As concluded above, the second-order equation (5.124) does only have zero or infinite asymptotic solutions. Thus, we will consider the function $y = x + 1$, where $x$ satisfies the second-order differential equation

$$\frac{d^2x}{dt^2} + 4\frac{dx}{dt} + 2x = 0. \tag{5.141}$$

Hence, the equation for $y$ reads

$$\frac{d^2y}{dt^2} + 4\frac{dy}{dt} + 2(y-1) = 0. \tag{5.142}$$

This equation has the desired property that $y$ levels off at one. It will be combined with the initial conditions $y(0) = 0.1$ and $dy/dt(0) = 1$. Figure 5.13a shows that the analytical solution of Eq. (5.142), which is given by $y = x + 1$ where $x$ is found in terms of Eq. (5.140), looks similar to the solution of the first-order differential equation shown in Fig. 5.12a. Figure 5.13b presents the normalized error $e/\Delta t$ of numerical solutions of Eq. (5.142), where $e$ is defined by Eq. (5.122). It turns out that the consideration of the second-order difference equation has a significant influence on the convergence behavior of the solution. The application of the discrete second-order difference equation (5.127) leads to a numerical instability for $\Delta t \leq 10^{-4}$. On the other hand, the system (5.131) of two discrete first-order equations provides a convergent solution

$$e = 0.4610\,\Delta t \tag{5.143}$$

for $\Delta t \leq 10^{-4}$. Actually, the choice $\Delta t \leq 10^{-3}$ is also sufficient to integrate the second-order differential equation accurately: $e/\Delta t = 0.4617$ at $\Delta t = 10^{-3}$ is very close to the asymptotic value $e/\Delta t = 0.4610$. Thus, the difference equation system (5.131) represents an accurate tool to solve Eq. (5.124) numerically.

## 5.6 Summary

Difference equations are used for two purposes: the modeling of observations given at points with a finite distance (e.g., population data that are available every 10 years) and for the solution of differential equations. Let us summarize the basic observations made in this chapter with regard to these two applications.

**Application to Modeling.** The evaluation of models that are obtained by the use of difference equations corresponds to the process applied to the evaluation of other models: see Sect. 1.4.3. On the other hand, the development of difference equation models faces some specific questions related to the type of difference equations considered, the modeling scenarios that can be covered in dependence on the range of model parameters considered, and the adjustment of model parameters to observations. Let us have a look at these closely related questions.

- First, which difference equation should be used for the modeling of any case? It is often reasonable to consider first a linear first-order equation provided there is basically one trend that has to be modeled (see, e.g., the trend shown in Fig. 5.1a). Usually, there are model deficiencies because a linear first-order equation does not offer a lot of flexibility. Such shortcomings can be addressed by using the more flexible concept of a second-order equation to modify the transition to an equilibrium state provided by a linear first-order equation (see, e.g., Fig. 5.4a). The use of nonlinear models is an efficient tool to describe competitive trends that affect changes of variables (see, e.g., the solution of the logistic population model shown in Fig. 5.10a: this curve reflects the competition of an exponential increase and an asymptotic trend to leveling off).
- Second, which range of variations of model parameters should be considered? This question is as relevant as the choice of the difference equation: we have to know which range of model parameters should be considered to simulate a desired or observed scenario. The answer to the latter question depends on the equation considered. There is not a big problem regarding the use of a linear first-order equation: the dependence of model features on the model parameters is obvious: see Fig. 5.1. Regarding the use of a linear second-order equation it needs an analysis of the dependence of eigenvalues on the model parameters (see the discussion of solution regimes of Samuelson's national income model). The understanding of basic solution features of a nonlinear equation requires an analysis of available equilibrium points and their stability.
- Third, how can we adjust model parameters in difference equations to obtain an optimal agreement with observations? There is usually no unique way to derive all the model parameters from any data set. A helpful way is the use of relations that ensure that the modeled equilibrium solution agrees with observations. For example, the setting of the equilibrium value $s = b / (1 - a)$ in a first-order equation or the setting of $s = C / (1 - A - B)$ in a second-order equation provides conditions for the model parameters. Other model parameters have to be determined such that the modeled transition to an asymptotic state optimally agrees with observed trends. An appropriate definition of model parameters supports the analysis of effects of parameter variations (as given, e.g., by the use of $\alpha$ and $\beta$ in Samuelson's model where $\alpha$ does fully control the equilibrium solution).

**Application to the Solution of Differential Equations.** The most relevant use of difference equations is their use for the formulation of differential equations, the numerical solution of differential equations, and the analytical solution of differential equations. Two examples for the application of difference equations to the solution of differential equations were presented in Sect. 5.5. Several other examples for the usefulness of difference equations will be given in Chap. 7 and Chap. 9: all the solutions of nonlinear equations presented in these chapters are obtained by solving the corresponding difference equations.

## 5.7 Exercises

**5.2.1** Consider the equation $y_n = -y_{n-1} + n$.

a) Use this equation to find $y_1$, $y_2$, $y_3$, $y_4$, $y_5$, and $y_6$ in dependence on the initial value $y_0$.

b) Use these expressions to guess the solution $y_n$, where $n = 0, 1, 2, \ldots$.

**5.2.2** Consider the equation $y_n = a\, y_{n-1} + b\, n + c$. Here, $a$, $b$, and $c$ are constants, and $n = 1, 2, \ldots$. The solution of this equation is given by

$$y_n = a^n\, y_0 + \frac{b}{1-a}\left[n - a\frac{1-a^n}{1-a}\right] + c\frac{1-a^n}{1-a}.$$

a) Show that this solution is correct for $n = 0$.

b) Show that this solution solves the equation $y_n = a\, y_{n-1} + b\, n + c$.

**5.2.3** Consider for the bracket term involved in exercise 5.2.2 the identity ($n = 0, 1, 2, \ldots$)

$$a^{n-1} + 2a^{n-2} + 3a^{n-3} + \cdots + (n-1)a^1 + n a^0 = \frac{1}{1-a}\left[n - a\frac{1-a^n}{1-a}\right].$$

a) Show the correctness of this identity for $n = 0$, $n = 1$, and $n = 2$.

b) Show the correctness of this identity for $n = 4$ and $a = 7$.

**5.2.4** Consider the difference equation $y_n = a\, y_{n-1} + b\, n + c$ and its solution given in exercise 5.2.2. Regarding the last term in the solution to this equation it is known that

$$\lim_{a\to 1}\frac{1-a^n}{1-a} = \lim_{a\to 1}\left(a^{n-1} + a^{n-2} + \cdots + a^1 + a^0\right) = n.$$

a) Calculate the limit $a \to 1$ of the identity given in exercise 5.2.3.

b) Find the solution of the equation $y_n = a\, y_{n-1} + b\, n + c$ for the case $a = 1$.

**5.2.5** Consider the equation $y_n = a\, y_{n-1} + b\, n + c$ and its solution given in exercise 5.2.2. Assume that $a$, $b$, $c$, and $y_0$ are nonzero.

a) Calculate the asymptotic solution features for $n \to \infty$ for the cases $a < 0$, $a = 0$, and $a > 0$.

b) There are observations that show an asymptotic behavior $y_n = 2\, e^{3n}$. Your task is to choose model parameters $a$, $b$, $c$, and $y_0$ such that one trend of the solution of the equation $y_n = a\, y_{n-1} + b\, n + c$ agrees with the observations. Explain the corresponding constraints for model parameters.

**5.2.6** Consider the compound interest formula $S_n = (1 + p\,/\,r)^n\, S_0$.

a) Consider two options to earn interest. A first option offers interest compounded monthly with an interest rate of 4%. A second option offers interest compounded annually with an interest rate of $100\,p$%. Which $p$ brings the same amount as given by the first option?

b) Calculate the interest rate that doubles any initial principal in five years. The conversion period is three months.

**5.2.7** Consider the simple interest formula $S_n = (1 + n\, p)\, S_0$ and the compound interest formula $S_n = (1 + p\,/\,r)^n\, S_0$. There are three options to earn interest. Company A offers simple interest at a rate of 6%. Company B offers compound interest at a 4% rate with a conversion period of one month. Company C offers compound interest at a 4% rate with a conversion period of three months.

a) Calculate for the three cases the amount on deposit after 5, 10, 15, and 20 years for any principal $S_0$.

b) Which interest offer maximizes the amount on deposit after 5, 10, 15, and 20 years?

**5.2.8** The loan on a house is \$200,000.

a) Calculate the monthly repayment needed to have the loan repaid after 30 years. The interest rate is 5%.

b) Calculate the total amount paid back on the loan.

**5.2.9** A company deposits a sum of money $S_0$ in a fund earning $100\,p$% interest compounded monthly. The company also deposits a sum $S_0$ in this fund at the end of each conversion period.

a) Find the difference equation for this problem and its solution.

b) Simplify the solution for the case that $p\,/\,r << 1$.

**5.2.10** A company deposits \$1000 in a fund earning 5% interest compounded monthly. In addition, the company also deposits a sum $D$ in this fund at the end of each conversion period.

a) Find the difference equation for this problem and its solution.
b) Which additional payment $D$ is needed to have a balance of \$30,000 in the account after three years?

**5.3.1** Consider the second-order difference equation ($n = 2, 3, \ldots$)

$$y_n = -\frac{1}{4} y_{n-2}.$$

The initial data are given by $y_0 = 2$ and $y_1 = 0$.
a) Find the solution of this difference equation.
b) Prove the validity of this solution.
c) Determine the asymptotic solution for $n \to \infty$.

**5.3.2** Find solutions for the following second-order difference equations, where $n = 2, 3, \ldots$. The initial data for all cases are given by $y_0 = 1$ and $y_1 = -1$.

*a*) $y_n = 2y_{n-2} - 1,$

*b*) $y_n = y_{n-1} - \frac{1}{4} y_{n-2} + \frac{1}{4},$

*c*) $y_n = 2y_{n-1} - 2y_{n-2} - 1.$

**5.3.3** Consider the second-order difference equation ($n = 2, 3, \ldots$)

$$y_n = 2y_{n-1} - y_{n-2} - 4.$$

The initial data are given by $y_0 = 1$ and $y_1 = 0$.
a) Find a particular solution $Y_n$ of this difference equation. Hint: consider the suitability of a quadratic function of $n$.
b) Rewrite the problem as an equation for $x_n = y_n - Y_n$. Hint: you have to provide initial data for $x_n$.
c) Find the solution $x_n$ that satisfies the equation and initial data for $x_n$.
d) Show that the resulting solution for $y_n$ satisfies both the equation for $y_n$ and the initial data for $y_n$.

**5.3.4** In 1202 Fibonacci, a famous Italian mathematician who is known for the spreading of the Hindu-Arabic numeral system in Europe, was interested in the reproduction of rabbits. He considered the following conditions:
- One male rabbit and one female rabbit have just been born.
- A rabbit will reach sexual maturity after one month.
- The gestation period of a rabbit is one month.
- A female rabbit gives birth every month after reaching sexual maturity.
- A female rabbit will always give birth to one male and one female rabbit.
- Rabbits never die.

a) Calculate the number of the pairs of rabbits for the first five months.
b) Derive the difference equation that describes the number of the pairs of rabbits per month.
c) Solve the difference equation.
d) Calculate how many pairs of rabbits will there be a year from now.

**5.3.5** Consider a modification of the red blood cell production model discussed in Sect. 5.3.4. In an attempt to formulate a better model, we assume that the number of red blood cells produced by the bone marrow is a constant $K$ (which means that $M_{n-1}$ in Eq. (5.80a) is replaced by $K$).
a) Solve the resulting difference equation analytically.
b) Assume that $R_0 = 0.9$, $f = 0.01$, and $K = 0.01$. Discuss the suitability of the resulting solution as a model for the red blood cell production.

**5.3.6** Consider the model $R_n = a\,R_{n-1} + b\,R_{n-2}$ for the number of red blood cells. Here, $a$ and $b$ are any model parameters (i.e., we do not consider $a = 1 - f$ and $b = f$ as in Sect. 5.3.4). We assume that the number $R_n$ of red blood cells can be described by the function $R_n = A\cos(\omega\, n - \delta)$, where $A$, $\omega$, and $\delta$ are model parameters.
a) Calculate the parameters $a$ and $b$ so that $R_n = A\cos(\omega\, n - \delta)$ is a solution of the second-order difference equation $R_n = a\,R_{n-1} + b\,R_{n-2}$. Hint: use the relation $\cos(z_1 + z_2) = \cos z_1 \cos z_2 - \sin z_1 \sin z_2$. You should also use the relations $\sin 2z = 2\sin z \cos z$ and $\cos 2z = 2\cos^2 z - 1$.
b) Use the results for $a$ and $b$ to explain the reason for the shortcomings of the expressions $a = 1 - f$ and $b = f$ used in Sect. 5.3.4.

**5.4.1** Consider the following nonlinear first-order difference equations, where $n = 1, 2, \ldots$. Find the equilibrium points and determine whether the equilibrium states are stable or not.

$$a)\ y_n = \frac{1}{1 + y_{n-1}},$$

$$b)\ y_n = \frac{y_{n-1}}{1 + y_{n-1}},$$

$$c)\ y_n = y_{n-1}(1 - y_{n-1}{}^2).$$

**5.4.2** The Ricker equation $y_n = \alpha y_{n-1} \exp(-\beta y_{n-1})$ is often used for the modeling of fish populations. In this equation, $\alpha$ is the maximal growth rate of the organism, and $\beta$ represents the inhibition of growth due to overpopulation (Edelstein-Keshet 2005).
a) Calculate the nonzero equilibrium solution to this equation.
b) Show under which conditions the equilibrium solution is stable.

**5.4.3** The Hassel model $y_n = \alpha y_{n-1}(1 + y_{n-1})^{-\beta}$, which may be seen as a limiting case of Ricker's model, has been used for the analysis of insect populations (Fulford et al. 1997). Here, $\alpha$ and $\beta$ are positive constant.

a) Calculate the nonzero equilibrium solution to this equation.

b) Show under which conditions the equilibrium solution is stable.

**5.5.1** Consider the logistic difference equation $y_n = y_{n-1} + a\, y_{n-1}(1 - y_{n-1}/K)$ that was considered in Sect. 5.4.2: see Eq. (5.97).

a) Determine the scaling of the model parameters with the time interval $\Delta t$ so that the discrete derivative $(y_n - y_{n-1})/\Delta t$ is independent of $\Delta t$.

b) Present the differential equation for the continuous function $y(t)$ that is related to the difference equation considered.

**5.5.2** Consider the equation $y_n = \alpha y_{n-1} \exp(-\beta y_{n-1})$ considered in exercise 5.4.2. Assume that $\alpha = e^A$, where $A$ is any parameter.

a) Find the scaling of the model parameters $A$ and $\beta$ with the time interval $\Delta t$ such that the discrete derivative $(y_n - y_{n-1})/\Delta t$ is independent of $\Delta t$ in the limit $\Delta t \to 0$.

b) Present the differential equation for the continuous function $y(t)$ that is related to the difference equation considered.

**5.5.3** Consider the population model $y_n = \alpha y_{n-1}(1 + y_{n-1})^{-\beta}$ considered in exercise 5.4.3. Assume that $\alpha = A^\beta$, where $A$ is any parameter.

a) Find the scaling of the model parameters $A$ and $\beta$ with the time interval $\Delta t$ such that the discrete derivative $(y_n - y_{n-1})/\Delta t$ is independent of $\Delta t$ in the limit $\Delta t \to 0$.

b) Present the differential equation for the continuous function $y(t)$ that is related to the difference equation considered.

# 6 Stochastic Changes

The characteristic features of solutions of several types of difference equations were discussed in Chap. 5 to see which sort of processes can be described in this way. These discussions did not account for randomness. However, randomness effects are relevant to many applications. Examples are given by the diffusion of substances in the atmosphere or water, the chaotic motion of molecules and other particles in fluids, and the development of population densities in time. Therefore, the deterministic methods presented in Chap. 5 will be extended in this chapter by the inclusion of randomness effects. The relevance of the stochastic difference equations considered in the following is that these equations provide the solutions of stochastic differential equations and equations for PDFs, which are the basic equations for the modeling of the evolution of stochastic processes. Hence, this chapter provides the basis for the discussions in Chaps. 8 and 10. In particular, the basic ingredient of stochastic evolution equations (the Wiener process) and the basic methodology for the solution of stochastic evolution equations (Monte Carlo simulation) will be introduced in this chapter.

This chapter is organized in analogy to the structure of Chap. 5. Section 6.1 motivates the consideration of randomness in difference equations for changes. Sections 6.2 and 6.3 introduce linear stochastic equations: their general solution features and applications to diffusion processes will be discussed. Section 6.4 then carries this on into linear equations for changes with delays. The discussion of the diffusion model in Sect. 6.3 will be extended in this way by the discussion of the more general Brownian motion model. This latter involves the basic features of models for molecular motion – a starting point for the modeling of fluids (see Chap. 10). Section 6.5 then carries the discussion of linear stochastic changes in Sects. 6.2 and 6.3 to the domain of nonlinear stochastic changes. This section generalizes the discussion of nonlinear population models in Sect. 5.4 by the inclusion of randomness. Section 6.6 summarizes the observations made in this chapter.

S. Heinz, *Mathematical Modeling*, DOI 10.1007/978-3-642-20311-4_6,

## 6.1 Motivation

**Diffusion.** Let us consider first an example to illustrate the need to analyze the properties of stochastic difference equations. Many (or even most) processes in nature and technology are driven by (temperature, energy, velocity, concentration, …) changes. Such processes are called diffusion (or dispersion) processes because the quantity considered (e.g., temperature) is distributed until an equilibrium state is established (i.e., until the differences that drive the process are minimized). There are many examples of diffusion processes. Diffusion is responsible for the distribution of sugar throughout a cup of coffee. Diffusion is the mechanism by which oxygen moves into our cells. Diffusion is of fundamental importance in many disciplines of physics, chemistry, and biology: diffusion is relevant to the sintering process (powder metallurgy, production of ceramics), the chemical reactor design, catalyst design in the chemical industry, doping during the production of semiconductors, and the transport of necessary materials such as amino acids within biological cells.

**Diffusion Model.** Let us consider a diffusion model to see the mathematical structure of such models. In particular, we consider the model ($n = 1, 2, \ldots$)

$$y_n = y_{n-1} + r\,\varepsilon_{n-1}. \tag{6.1}$$

Here, $y_n$ refers to the position of any particle (for example, the height of any tracer above ground). The initial position $y_0$ is assumed to be given. For simplicity, we assume that $y_0$ is a deterministic parameter such that all the particles start at the same position. The random number $\varepsilon_k$ accounts for the effect of randomness (it models random up and down motions due to diffusion), and the parameter $r$ is a deterministic parameter that modifies the intensity of randomness. The quantity $\varepsilon_k$ is assumed to be normally distributed. The specification of a normal PDF requires the definition of the mean and variance. Here, we assume that $\varepsilon_k$ is a standardized normal process with zero mean $<\varepsilon_k> = 0$ and variance $<\varepsilon_k^2> = 1$. These settings specify $\varepsilon_k$ at any $k$. In this way, nothing is said regarding the relation of $\varepsilon_k$ and $\varepsilon_m$, where $k \neq m$. At every step, the noise process is considered to provide a new input that is independent of previous noise values. Therefore, we assume that $\varepsilon_k$ and $\varepsilon_m$ are independent random variables, $<\varepsilon_k\,\varepsilon_m> = 0$, for $k \neq m$. These assumptions for $\varepsilon_k$ can be summarized in the following way:

$$\left\langle \varepsilon_k \right\rangle = 0, \qquad \left\langle \varepsilon_k \varepsilon_m \right\rangle = \delta_{km}. \tag{6.2}$$

Here, $\delta_{km}$ refers to the Kronecker delta that is defined by the conditions $\delta_{km} = 1$ for $k = m$ and $\delta_{km} = 0$ for $k \neq m$. The solution to the stochastic difference equation (6.1) can be found by having a look at the implications of this difference equation.

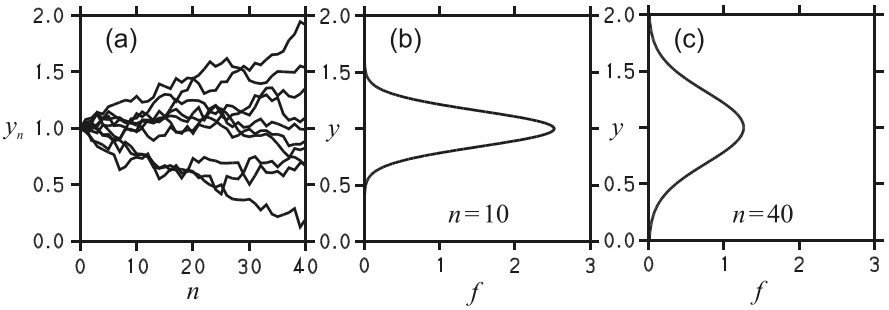



**Fig. 6.1.** An illustration of features of the diffusion model (6.1). Ten realizations of solutions are shown in (a), where $n$ refers to the step number. The PDF $f(y)$ of position values is shown in (b) and (c) at $n = 10$ and $n = 40$, respectively.

Regarding the first three values $y_n$ we find

$$y_1 = y_0 + r\,\varepsilon_0, \tag{6.3a}$$

$$y_2 = y_1 + r\,\varepsilon_1 = y_0 + r\left(\varepsilon_0 + \varepsilon_1\right), \tag{6.3b}$$

$$y_3 = y_2 + r\,\varepsilon_2 = y_0 + r\left(\varepsilon_0 + \varepsilon_1 + \varepsilon_2\right). \tag{6.3c}$$

Here, we substituted $y_1 = y_0 + r\,\varepsilon_0$ in $y_2$, and $y_2 = y_0 + r\,(\varepsilon_0 + \varepsilon_1)$ in $y_3$. Therefore, we can conclude that the solution $y_n$ is given by ($n = 0, 1, \ldots$)

$$y_n = y_0 + r\left(\varepsilon_0 + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{n-1}\right). \tag{6.4}$$

Usually, a problem is solved if the solution for any equation is found: the solution can be illustrated and the effect of parameter variations can be studied. Examples for the solution (6.4) are shown in Fig. 6.1, where $y_0 = 1$ and $r = 0.05$ are applied. One realization is obtained by using one specific choice of random numbers in Eq. (6.4). It is evident that every realization of Eq. (6.4) is different: it is impossible to analyze the solution features in this way. The only way to analyze such solutions is to look at the statistics of this process at every $n$. In particular, the evolution of the PDF has to be determined in dependence on $n$. Examples for PDFs of position values are given in Figs. 6.1b–c. The PDFs that are shown in these figures represent normal PDFs with mean one and variance $n\,r^2$ (a detailed discussion of the properties of diffusion models is provided in Sect. 6.3). These PDFs reveal the nature of a diffusion process: the variance increases linearly with $n$, which simulates the distribution of the tracer considered in space.

**Processes Considered.** There is a variety of stochastic processes, and it would be hardly possible to cover all these processes here. Therefore, we will focus in the following on processes with relatively little stochastic changes (i.e., diffusion processes). Such processes are relevant to a huge variety of processes, but there

are also cases for which the use of other concepts is more appropriate. One example for such a case is given by processes that jump from one integer to another integer (e.g., a population density that changes like 79, 81, 94, 103). Another example for such a case is given by processes that are bounded, for example by zero and one, according to their definition (e.g., mass fractions of chemical species). The modeling of such processes in terms of a normally distributed noise can imply unphysical negative values of variables.

**Questions Considered.** The illustration of the modeling of diffusion processes in terms of Eq. (6.1) leads to questions like the following ones:

- What is an appropriate noise model, and what are its characteristic properties?
- How does a noise model affect the statistical properties of a process considered?
- How can we evaluate the suitability of a model for a stochastic process?

These and other questions will be discussed in the following sections with regard to several relevant diffusion processes in mechanics and population ecology.

## 6.2 Linear Stochastic Changes

Let us discuss first an extension of Eq. (6.1): we consider for a variable $y_n$ the first-order linear difference equation ($n = 1, 2, \ldots$)

$$y_n = a\, y_{n-1} + b + r\, \varepsilon_{n-1}. \tag{6.5}$$

Here, $a$, $b$, and $r$ are any deterministic model parameters. We assume that $y_n$ has a known normally distributed initial value $y_0$ at $n = 0$. The noise term $\varepsilon_k$ is assumed to be normally distributed with mean zero and variance one. For $k \neq m$, $\varepsilon_k$ and $\varepsilon_m$ are independent random variables. Therefore, $\varepsilon_k$ is characterized by the properties

$$\langle \varepsilon_k \rangle = 0, \qquad \langle \varepsilon_k \varepsilon_m \rangle = \delta_{km}. \tag{6.6}$$

It should be mentioned that these assumptions are sufficiently general: any nonzero mean value of $\varepsilon_k$ can be combined with $b$, and any variance of $\varepsilon_k$ unequal to one can be combined with $r$. We also have to specify the correlation between $\varepsilon_k$ and $y_0$ because $y_0$ is a random variable. The noise model should be independent of the initial distribution. Thus, we assume here that $\varepsilon_k$ and $y_0$ represent independent random variables. In the following we will derive and illustrate first the analytical consequences of Eq. (6.5) for the one-point statistics (the properties of $y_k$ at one $k$) and correlations (between values $y_k$ and $y_m$ with $k \neq m$). Second, we will consider the typical features of numerical solutions of Eq. (6.5), i.e., we discuss the properties of Monte Carlo simulations. The results obtained in this section will be used in the following sections of this chapter.

### 6.2.1 One-Point Statistics

**Solution.** In accord with the derivation of the solution of a deterministic linear first-order difference equation in Chap. 5, let us consider first the implications of Eq. (6.5) for $n = 1$, 2, and 3,

$$y_1 = a\,y_0 + b + r\,\varepsilon_0, \tag{6.7a}$$

$$\begin{aligned} y_2 &= a\,y_1 + b + r\,\varepsilon_1 = a(a\,y_0 + b + r\,\varepsilon_0) + b + r\,\varepsilon_1 \\ &= a^2 y_0 + a(b + r\,\varepsilon_0) + b + r\,\varepsilon_1, \end{aligned} \tag{6.7b}$$

$$\begin{aligned} y_3 &= a\,y_2 + b + r\,\varepsilon_2 = a(a^2 y_0 + a(b + r\,\varepsilon_0) + b + r\,\varepsilon_1) + b + r\,\varepsilon_2 \\ &= a^3 y_0 + a^2(b + r\,\varepsilon_0) + a(b + r\,\varepsilon_1) + b + r\,\varepsilon_2. \end{aligned} \tag{6.7c}$$

Here, the expressions for previous values of $y_n$ are used for representing the next $y_n$ in terms of $y_0$. By following these patterns we can find the solution $y_n$ of Eq. (6.5) for $n = 1, 2, \ldots,$

$$y_n = a^n y_0 + a^{n-1}(b + r\,\varepsilon_0) + \cdots + a^1(b + r\,\varepsilon_{n-2}) + a^0(b + r\,\varepsilon_{n-1}). \tag{6.8}$$

This solution generalizes the solution (6.4), which is obtained for $a = 1$ and $b = 0$, and it generalizes the solution (5.9), which is obtained by neglecting $\varepsilon_k$. As illustrated in the introduction, the knowledge of such a random analytical solution is not directly helpful. However, this analytical solution can be used for deriving statistics of the stochastic process considered.

**Moments.** First, let us have a look at low-order moments of $y_n$ that are implied by Eq. (6.8). The mean value $\langle y_n \rangle$ can be obtained by averaging Eq. (6.8). Due to the fact that $\varepsilon_k$ has a zero mean, we obtain

$$\langle y_n \rangle = a^n \langle y_0 \rangle + b(a^{n-1} + \cdots + a^1 + a^0). \tag{6.9}$$

This expression is equal to the solution (5.9) of the corresponding deterministic difference equation. We use the identity (5.10),

$$a^{n-1} + a^{n-2} + \cdots + a^1 + a^0 = \frac{1 - a^n}{1 - a}, \tag{6.10}$$

to write the mean value more conveniently,

$$\langle y_n \rangle = a^n \langle y_0 \rangle + b\frac{1 - a^n}{1 - a}. \tag{6.11}$$

To derive a corresponding expression for the variance it is helpful to use Eq. (6.9) for deriving an expression for the fluctuating solution $\tilde{y}_n = y_n - \langle y_n \rangle$. By taking the difference between Eqs. (6.8) and (6.9) we obtain

$$\tilde{y}_n = a^n\,\tilde{y}_0 + r(a^{n-1}\,\varepsilon_0 + a^{n-2}\,\varepsilon_1 + \cdots + a^1\,\varepsilon_{n-2} + a^0\,\varepsilon_{n-1}). \tag{6.12}$$

The consistency of this expression may be seen by averaging it – which results in zero on both sides. Equation (6.12) can be used for calculating the variance of $y_n$. By squaring Eq. (6.12) and taking the average we find

$$\left\langle \tilde{y}_n^{\,2} \right\rangle = a^{2n} \left\langle \tilde{y}_0^{\,2} \right\rangle + r^2 (a^{2(n-1)} + a^{2(n-2)} + \cdots + a^2 + a^0). \tag{6.13}$$

The latter expression makes use of the fact that $\varepsilon_k$ and $\varepsilon_m$ are independent random variables for $k \neq m$, and that $\varepsilon_k$ and $y_0$ are also independent. Hence, only averaged squared terms of the contributions in Eq. (6.12) contribute to the variance formula (6.13). Equation (6.13) can be rewritten in correspondence to Eq. (6.11) for the mean by using Eq. (6.10) for $a^2$,

$$(a^2)^{n-1} + (a^2)^{n-2} + \cdots + (a^2)^1 + (a^2)^0 = \frac{1 - a^{2n}}{1 - a^2}. \tag{6.14}$$

Correspondingly, we obtain

$$\left\langle \tilde{y}_n^{\,2} \right\rangle = a^{2n} \left\langle \tilde{y}_0^{\,2} \right\rangle + r^2 \frac{1 - a^{2n}}{1 - a^2}. \tag{6.15}$$

Higher-order moments can be calculated in the same way. Due to the property of $y_0$ and $\varepsilon_k$ to be normally distributed it is not hard to see that the odd moments of $\tilde{y}_n$ disappear. The calculation of even moments of $\tilde{y}_n$ in this way is relatively complicated.

**PDF.** A convenient way to completely determine the one-point statistics implied by Eq. (6.5) is to calculate the PDF of $y_n$. The latter can be done on the basis of the theorem (10.75),

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^{\,2}) \quad \Rightarrow \quad \sum_{i=1}^{N} X_i \sim \mathcal{N}\left( \sum_{i=1}^{N} \mu_i, \sum_{i=1}^{N} \sigma_i^{\,2} \right). \tag{6.16}$$

This theorem states the following: We consider random numbers $X_i$ ($i = 1, N$) that are normally distributed with mean $\mu_i$ and variance $\sigma_i^2$. Then, the sum of $X_i$ values is also normally distributed. The mean of this normal PDF is the sum of all $\mu_i$, and the variance of this normal PDF is the sum of all variances $\sigma_i^2$. For our case, the sum of $X_i$ values is given by the terms in Eq. (6.8). Due to the properties of $y_0$ and $\varepsilon_k$, all single contributions are normally distributed. Thus, the PDF of $y_n$ also has to be normally distributed. The mean of the PDF of $y_n$ is given by Eq. (6.9), which represents the sum of the means of all contributions. The variance of the PDF of $y_n$ is given by Eq. (6.13), which represents the sum of the variances of all contributions. These observations are summarized by the formula

$$f_n(y) = \frac{1}{\sqrt{2\pi \left\langle \tilde{y}_n^{\,2} \right\rangle}} \exp\left\{ - \frac{\left( y - \left\langle y_n \right\rangle \right)^2}{2 \left\langle \tilde{y}_n^{\,2} \right\rangle} \right\} \tag{6.17}$$

for the PDF $f_n(y)$ of $y_n$. Here, $y$ is the sample space variable corresponding to the random variable $y_n$. The mean and variance of this PDF are given by Eqs. (6.11) and (6.15), respectively. The notation $f_n(y)$ refers to the dependence of $n$, this means the PDF (6.17) evolves with $n$. Equation (6.17) can be applied for finding higher-order moments by using the properties of normal distributions discussed in Sect. 4.3.2.

### *6.2.2 Correlations*

**Correlation Relevance.** A basic characterization of a stochastic state is given by the mean and standard deviation, which explain the typical value and range of variations of any random variable considered. The consideration of the evolution of a stochastic process leads to the additional question about the typical lifetime of fluctuations: to model the dynamics of fluctuations in terms of a stochastic model we need to know after which characteristic time (one nanosecond, or one second, or one hour?) fluctuations will disappear. Information about the typical lifetime of fluctuations is available via the correlation $<\tilde{y}_n \tilde{y}_{n+m}>$. This quantity is considered at a fixed $n$ as a function of $m = 0, 1, 2, \ldots$. Usually, we consider the normalized correlation function

$$C_n(m) = \frac{\langle \tilde{y}_n \tilde{y}_{n+m} \rangle}{\langle \tilde{y}_n \tilde{y}_n \rangle}, \tag{6.18}$$

which is equal to the correlation coefficient between $y_n$ and $y_m$ if the variance is stationary (if $<\tilde{y}_n \tilde{y}_n> = <\tilde{y}_{n+m} \tilde{y}_{n+m}>$). It is often found that $C_n(m)$ is a decreasing function of $m$. By analyzing at which $m$ the correlation function $C_n(m)$ is significantly reduced (e.g., below 1%), we know the typical lifetime of fluctuations.

**Correlation Calculation.** Let us calculate the correlation $<\tilde{y}_n \tilde{y}_{n+m}>$ for the process $y_n$ determined by Eq. (6.5). The simplest way to perform this calculation is to consider the pattern of these correlations. This approach applies the difference equation for fluctuations, which is obtained by taking the difference between Eq. (6.5) for $y_n$ and the implied equation for the mean value $<y_n>$,

$$\tilde{y}_n = y_n - \langle y_n \rangle = a\, y_{n-1} + b + r\, \varepsilon_{n-1} - \left(a \langle y_{n-1} \rangle + b\right) = a\, \tilde{y}_{n-1} + r\, \varepsilon_{n-1}. \tag{6.19}$$

We consider this $\tilde{y}_n$ at $n+1$, multiply it by $\tilde{y}_n$ and average this expression,

$$\langle \tilde{y}_n \tilde{y}_{n+1} \rangle = \langle \tilde{y}_n (a\, \tilde{y}_n + r\, \varepsilon_n) \rangle = a \langle \tilde{y}_n \tilde{y}_n \rangle + r \langle \tilde{y}_n \varepsilon_n \rangle. \tag{6.20}$$

The last term here can be evaluated by using Eq. (6.12) for $\tilde{y}_n$. Due to the fact that $\varepsilon_n$ is independent of all its previous values and $y_0$, we find that the last term in

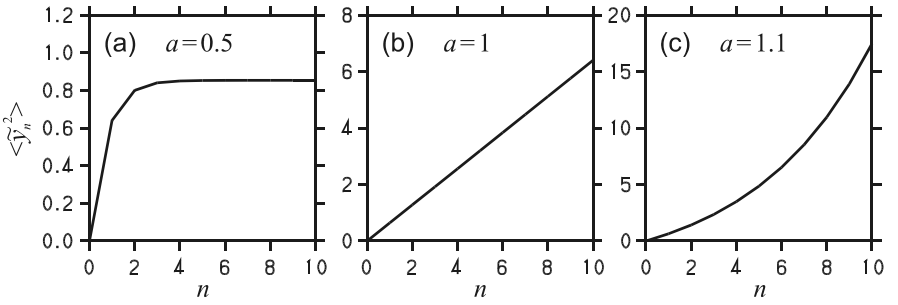



**Fig. 6.2.** An illustration of the variance (6.15) in dependence on $n$. Here, $y_0 = 0$ and $r = 0.8$ are applied. The model parameter $a$ applied is given in the plots.

Eq. (6.20) has to disappear, $<\tilde{y}_n\varepsilon_n> = 0$. Hence, $<\tilde{y}_n\tilde{y}_{n+1}>$ is given by

$$\langle \tilde{y}_n \tilde{y}_{n+1} \rangle = a \langle \tilde{y}_n \tilde{y}_n \rangle. \tag{6.21}$$

The corresponding features for $<\tilde{y}_n\tilde{y}_{n+2}>$ can be found correspondingly. First, we use Eq. (6.19) for the representation

$$\langle \tilde{y}_n \tilde{y}_{n+2} \rangle = \langle \tilde{y}_n (a\,\tilde{y}_{n+1} + r\,\varepsilon_{n+1}) \rangle = a \langle \tilde{y}_n \tilde{y}_{n+1} \rangle + r \langle \tilde{y}_n \varepsilon_{n+1} \rangle. \tag{6.22}$$

The last term has to disappear: the reasoning is the same as that used for showing that $<\tilde{y}_n\varepsilon_n> = 0$. By means of Eq. (6.21) we find

$$\langle \tilde{y}_n \tilde{y}_{n+2} \rangle = a^2 \langle \tilde{y}_n \tilde{y}_n \rangle. \tag{6.23}$$

By using the same approach for deriving corresponding expressions for $<\tilde{y}_n\tilde{y}_{n+3}>$ and so on, we find the general expression for correlations to be given by

$$\langle \tilde{y}_n \tilde{y}_{n+m} \rangle = a^m \langle \tilde{y}_n \tilde{y}_n \rangle, \tag{6.24}$$

where $m = 0, 1, 2, \ldots$. The setting $m = 0$ results in an identity for the variance of $y_n$. The normalized correlation function $C_n(m)$ defined by Eq. (6.18) is given by

$$C_n(m) = a^m. \tag{6.25}$$

It should be pointed out that $C_n(m)$ is independent of $n$ for the process considered.

### *6.2.3 Solution Features*

Let us illustrate the typical statistical properties of solutions to the difference equation (6.5). We will consider here the means and variances that determine the normal distribution (6.17) and the normalized correlation function $C_n(m)$.

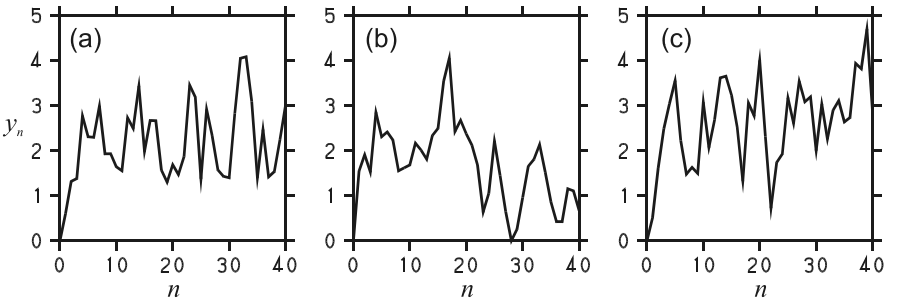



**Fig. 6.3.** An illustration of realizations of solutions of the difference equation (6.26) in dependence on $n$. Here, $a = 0.5$, $b = 1$, $r = 0.8$, and $y_0 = 0$ are applied.

**Means and Variances.** Equation (6.11) for $\langle y_n \rangle$ agrees with the solution (5.15) of the corresponding deterministic equation (5.6) where $\varepsilon_k = 0$. Therefore, $\langle y_n \rangle$ is characterized by the features of the deterministic $y_n$ that are illustrated in Fig. 5.1. Figure 6.2 illustrates that the variance shows a very similar behavior as the mean $\langle y_n \rangle$ (there are no curves corresponding to the curves with negative values of $a$ in Fig. 5.1 because the variance depends on $a^2$). The reason for that is the very similar structure of formulas: instead of $a$, $b$, and $\langle y_0 \rangle$ in the mean formula (6.11) we have $a^2$, $r^2$ and $\langle \tilde{y}_0^{\,2} \rangle$ in the variance formula (6.15). Regarding the modeling of processes, the most relevant case is given if $|a| < 1$ such that an equilibrium state is achieved. The corresponding equilibrium values (for large values of $n$) of means (6.11) and variances (6.15) are given by $\langle y_e \rangle = b / (1 - a)$ and $\langle \tilde{y}_e^{\,2} \rangle = r^2 / (1 - a^2)$. These relations can be used for the determination of the model parameters $b$ and $r^2$ by using known equilibrium values for the mean and variance. The model parameter $a$ has to be chosen such that the transition from the initial value to the equilibrium value is represented as good as possible. We note that the equilibrium variance is larger than the variance $r^2$ of the noise process $r\,\varepsilon_k$.

**Correlations.** The normalized correlation function (6.25) reveals the following features: First, usually it does not make sense to consider negative parameter $a$ values: $C_n(m)$ will oscillate then similar as the mean values in the lower plots of Fig. 5.1. The modeling of correlations in this way is not helpful for applications. For values $a > 1$, $C_n(m)$ will increase. Such a behavior is unreasonable because there is usually no mechanism that can increase correlations. For $a = 1$ we find $C_n(m) = 1$, which means that the correlation will never change. Such a case was considered in the introduction: see Eq. (6.1). Such a correlation behavior is helpful for the modeling of real processes if there is no mechanism that destroys correlations. The last case is given if $0 < a < 1$. The correlation function $C_n(m)$ will decay with $n$ in this case. This is the common behavior of correlations: memory is usually lost. This discussion reveals that the correlation function shows a meaningful behavior if the condition $0 < a \leq 1$ is satisfied.

### 6.2.4 Monte Carlo Simulation

**Example.** The results obtained in Sects. 6.2.1 and 6.2.2 lead to the question of how such conclusions can be validated. One way of doing this is to confirm such findings on the basis of simulations, i.e., by an analysis of simulation results. Analytical conclusions as those obtained in Sects. 6.2.1 and 6.2.2 are often unavailable. In this case, the only way of obtaining insight into the statistics of processes is the analysis of simulation results. The process of doing this will be illustrated in this section. We consider again Eq. (6.5),

$$y_n = a\, y_{n-1} + b + r\, \varepsilon_{n-1}, \tag{6.26}$$

but we work now with specified model parameter values: we assume $a = 0.5$, $b = 1$, $r = 0.8$, and $y_0 = 0$. Several realizations of Eq. (6.26) are illustrated in Fig. 6.3. Equation (6.26) is only used here for the generation of simulation results that can be analyzed. The analysis of the evolution of the mean and variance, the PDF and normalized correlation function will be described in the following.

**One-Point Statistics.** The mean $<y_n>$, standard deviation $<\tilde{y}_n^{\,2}>^{1/2}$ and PDF $f_n(y)$ can be calculated at every $n$ as described in Sect. 4.4.1. The numerical results are shown in Fig. 6.4 in comparison to the corresponding analytical results presented in Sect. 6.2.1. It may be seen that there is no observable difference between the numerical and theoretical results for the $10^7$ realizations applied (the effect of variations of the number of realizations was already described in Sect. 4.4.1). The evolution of the PDF with $n$ is illustrated by two examples: the PDF $f_1(y)$ at $n = 1$ and the PDF $f_{10}(y)$ at $n = 10$. In consistency with the evolution of $<y_n>$ shown in Fig. 6.4a, Figs. 6.4c–d show the variation of the mean value $<y_n>$ (which is equal here to the PDF maximum position) with $n$: we have $<y_1> = 1$ and $<y_{10}> = 2$. In correspondence to Fig. 6.4b, the standard deviation increases with $n$: we have $<\tilde{y}_1^{\,2}>^{1/2} = 0.80$ and $<\tilde{y}_{10}^{\,2}>^{1/2} = 0.92$.

**Correlations.** The normalized correlation function $C_n(m)$ can be calculated numerically in the following way ($N$ refers to the number of realizations applied and $j = 1, N$):

1. Use the difference equation (6.26) to find $y_n(j)$, where $n$ is a fixed value. Store:

$$\bullet\ \langle y_n \rangle \qquad \bullet\ \tilde{y}_n(j) = y_n(j) - \langle y_n \rangle \qquad \bullet\ \langle \tilde{y}_n^{\,2} \rangle \tag{6.27a}$$

2. Continue with the solution of the difference equation (6.26) to find $y_{n+m}(j)$ for varying $m$ values. Calculate for each $m = 0, 1, 2, \ldots$:

$$\bullet\ \langle y_{n+m} \rangle \qquad \bullet\ \tilde{y}_{n+m}(j) = y_{n+m}(j) - \langle y_{n+m} \rangle$$

$$\bullet\ C_n(m) = \frac{\langle \tilde{y}_n \tilde{y}_{n+m} \rangle}{\langle \tilde{y}_n \tilde{y}_n \rangle} = \frac{1}{\langle \tilde{y}_n^{\,2} \rangle} \frac{1}{N} \sum_{j=1}^{N} \tilde{y}_n(j)\, \tilde{y}_{n+m}(j) \tag{6.27b}$$

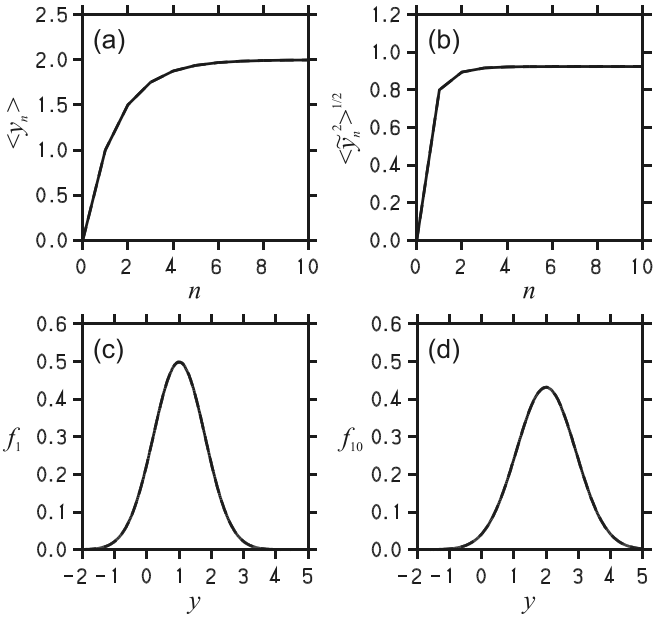



**Fig. 6.4.** The analysis of Monte Carlo simulation results. The *solid lines* in (a) and (b) show the mean value and standard deviation of $y_n$ obtained by Monte Carlo simulation in dependence on $n$. The *solid lines* in (c) and (d) show the PDF $f_1$ at $n = 1$ and the PDF $f_{10}$ at $n = 10$ obtained by Monte Carlo simulation, respectively. $10^7$ realizations and a filter width $\Delta y = 0.1$ were applied to calculate these PDFs. The corresponding analytical results are shown by *dashed lines* in all these figures (there is no observable difference).

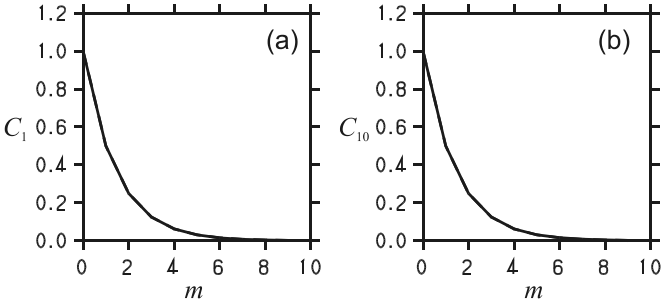



**Fig. 6.5.** The analysis of Monte Carlo simulation results. The *solid lines* in (a) and (b) show the normalized correlation functions $C_1$ at $n = 1$ and $C_{10}$ at $n = 10$ in dependence on $m$, respectively. $10^7$ realizations were applied for this calculation. The corresponding theoretical result (6.25) is given in both plots by a *dashed line* (there is no observable difference).

The calculation of the normalized correlation function $C_n(m)$ has to be performed carefully. For example, the fluctuations $\tilde{y}_n(j)$ and $\tilde{y}_{n+m}(j)$ have to be calculated in terms of the correct mean values $<y_n>$ and $<y_{n+m}>$, respectively. It is relevant that the correlation function is calculated as mean value of $\tilde{y}_n(j)\ \tilde{y}_{n+m}(j)$, where $\tilde{y}_n(j)$ and $\tilde{y}_{n+m}(j)$ denote the values of one realization at different positions. The correlation functions $C_1(m)$ at $n = 1$ and $C_{10}(m)$ at $n = 10$ are shown in Fig. 6.5 in comparison to the corresponding analytical results. As found with regard to the one-point statistics, there is no observable difference between the numerical and analytical results. The conclusion of Eq. (6.25) that $C_n(m)$ is independent of $n$ is also confirmed by Fig. 6.5: there is no observable difference between the $C_n(m)$ curves in both plots.

## 6.3 Diffusion

Let us consider random walk models (which are also called drunkard's walk models) to illustrate the application of the linear stochastic first-order difference equation discussed in Sect. 6.2. A random walk refers to a trajectory that results from taking successive random steps. Random walk models represent the basis for diffusion models that are applied in many areas (Durbin 1983): see Sect. 6.3.3. From a mathematical point of view, the focus of this section is on the introduction of the basic ingredient of stochastic evolution equations: the Wiener process. In the following we will analyze the features of the random walk model (6.1),

$$y_n = y_{n-1} + r\,\varepsilon_{n-1}. \tag{6.28}$$

Here, $y_n$ is seen as a position. The random variable $\varepsilon_{n-1}$ is normality distributed and characterized by $<\varepsilon_k> = 0$ and $<\varepsilon_k\ \varepsilon_m> = \delta_{km}$. We will also assume that $\varepsilon_k$ is independent of the random initial position $y_0$.

### *6.3.1 Random Walk Model*

**One-Point Statistics.** The properties of solutions to Eq. (6.28) can be seen by specifying the results for the more general model (6.5): we have to set $a = 1$ and $b = 0$. The discussion in Sect. 6.2 showed that $y_n$ is normally distributed. According to Eq. (6.9), the mean of $y_n$ equals the mean initial value for the case $a = 1$ and $b = 0$ considered,

$$\langle y_n \rangle = \langle y_0 \rangle. \tag{6.29}$$

Equation (6.13) shows that the variance of $y_n$ is given by

$$\left\langle \widetilde{y}_n{}^2 \right\rangle = \left\langle \widetilde{y}_0{}^2 \right\rangle + n r^2 . \tag{6.30}$$

Hence, $y_n$ is normally distributed according to

$$y_n \sim \mathcal{N}\left( \langle y_0 \rangle, \left\langle \widetilde{y}_0{}^2 \right\rangle + n r^2 \right). \tag{6.31}$$

The corresponding features of statistics are illustrated in Fig. 6.1: the mean value is constant, and the variance increases with $n$.

**Correlations.** The correlations of $y_n$ are determined according to Eq. (6.25) by

$$\left\langle \widetilde{y}_n \widetilde{y}_{n+m} \right\rangle = \left\langle \widetilde{y}_n \widetilde{y}_n \right\rangle, \tag{6.32}$$

where $m = 0, 1, 2, \ldots$. The meaning of this result can be seen in a better way by writing this relation

$$\left\langle \widetilde{y}_n (\widetilde{y}_{n+m} - \widetilde{y}_n) \right\rangle = 0. \tag{6.33}$$

This means that the change $\widetilde{y}_{n+m} - \widetilde{y}_n$ is uncorrelated to $\widetilde{y}_n$. The validity of this conclusion can be seen by proving that $\widetilde{y}_{n+m} - \widetilde{y}_n$ is caused by noise contributions that are uncorrelated to $\widetilde{y}_n$ (see exercise 6.3.1). The steady addition of uncorrelated noise contributions then implies the steady increase of the variance of $y_n$.

**Time Dependence.** A closer look at the variance expression (6.30) reveals the following. We may use Eq. (6.28) as a model for a continuous diffusion in time $n\,\Delta t$. For this case, the variance (6.30) should be a function of $n\,\Delta t$. The latter is the case if we parameterize the noise coefficient $r$ by

$$r = \sqrt{D\,\Delta t}\,. \tag{6.34}$$

It is relevant to note that we do not have to assume here and in the following that $\Delta t \to 0$. The model parameter $D$ represents a non-negative variable that is called the diffusion coefficient. In terms of the relation $r = (D\,\Delta t)^{1/2}$ we can write the variance $< \widetilde{y}_n{}^2 > = < \widetilde{y}_0{}^2 > + D\,n\,\Delta t$. Therefore, $D$ is the derivative of the variance by time $n\,\Delta t$. This fact explains why $D$ is called the diffusion coefficient: $D$ determines the increase of the position variance (which describes the spreading of a plume). By using $r = (D\,\Delta t)^{1/2}$, we can write the diffusion model (6.28) as

$$y_n = y_{n-1} + \sqrt{D\,\Delta t}\;\varepsilon_{n-1}. \tag{6.35}$$

According to (6.31), $y_n$ is normally distributed,

$$y_n \sim \mathcal{N}\left( \langle y_0 \rangle, \left\langle \widetilde{y}_0{}^2 \right\rangle + D\,n\,\Delta t \right). \tag{6.36}$$

Equation (6.32) implies for the normalized correlation function $C_n(m)$ that

$$C_n(m) = 1. \tag{6.37}$$

### 6.3.2 The Wiener Process

**Model Reformulation.** The noise term in Eq. (6.35) involves three components: the random variable $\varepsilon_{n-1}$, the scaling parameter $(\Delta t)^{1/2}$, and the diffusion coefficient $D$. The diffusion coefficient will change with the problem considered, but the first two components are the same for many kinds of problems. Thus, it is helpful to split the writing of the noise term in Eq. (6.35) into one contribution that represents the properly scaled noise, and another contribution given by the case-dependent diffusion coefficient $D$. The latter can be achieved by defining a variable $W_n$ by the difference equation of $y_n$ where $D = 1$,

$$W_n = W_{n-1} + \sqrt{\Delta t}\, \varepsilon_{n-1}. \tag{6.38}$$

To simplify this relation we introduce the abbreviation $\Delta W_{n-1} = W_n - W_{n-1}$, such that Eq. (6.38) reads

$$\Delta W_{n-1} = \sqrt{\Delta t}\, \varepsilon_{n-1}. \tag{6.39}$$

By making use of this definition of $\Delta W_{n-1}$ we can write the model (6.35) as

$$y_n = y_{n-1} + \sqrt{D}\, \Delta W_{n-1}. \tag{6.40}$$

This writing represents the standard formulation of stochastic difference and differential equations (see Chaps. 8 and 10). The noise model is written in terms of the case-independent scaled random variable $\Delta W_{n-1}$, which enables the calculation of statistics that are functions of time $t = n\, \Delta t$, and the case-dependent diffusion coefficient $D$.

**Wiener Process.** The stochastic process $W_n$ is called the Wiener process. It represents the most important stochastic process. Thus, let us have a closer look at its properties. The comparison of Eq. (6.38) for $W_n$ with Eq. (6.35) for $y_n$ shows that statistical properties of $W_n$ can be obtained by setting $D = 1$ in the corresponding statistics of $y_n$. Therefore, we find that the Wiener process $W_n$ is normally distributed according to

$$W_n \sim \mathcal{N}\left(\langle W_0 \rangle, \left\langle \widetilde{W}_0^{\,2} \right\rangle + n\, \Delta t\right). \tag{6.41}$$

As found for $y_n$, the correlations of $W_n$ are determined by

$$\left\langle \widetilde{W}_n \widetilde{W}_{n+m} \right\rangle = \left\langle \widetilde{W}_n \widetilde{W}_n \right\rangle. \tag{6.42}$$

**Wiener Process Changes.** In addition to the properties of $W_n$, it is helpful for the discussions below to consider the properties of the change $\Delta W_n$ of a Wiener process. The definition $\Delta W_n = (\Delta t)^{1/2}\, \varepsilon_n$ shows that $\Delta W_n$ is normally distributed,

$$\Delta W_n \sim \mathcal{N}(0,\, \Delta t). \tag{6.43}$$

The correlation properties of $\Delta W_n$ also follow from $\Delta W_n = (\Delta t)^{1/2}\, \varepsilon_n$,

$$\left\langle \Delta W_n \Delta W_m \right\rangle = \begin{cases} \Delta t & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases}, \tag{6.44}$$

where we made use of $<\varepsilon_k\, \varepsilon_m> = \delta_{km}$. We applied $\Delta \widetilde{W}_n = \Delta W_n$, which is the same because of $<\Delta W_n> = 0$. By dividing both sides by $(\Delta t)^2$ we can write

$$\left\langle \frac{\Delta W_n}{\Delta t} \frac{\Delta W_m}{\Delta t} \right\rangle = \begin{cases} \dfrac{1}{\Delta t} & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases}. \tag{6.45}$$

This result means that the variance of the derivative $\Delta W_n / \Delta t$ of $W_n$ does not exist for $\Delta t \to 0$: the variance goes to infinity. Consequently, $W_n$ is not differentiable because the probability for the appearance of values of $\Delta W_n / \Delta t$ that are larger than any limit is equal to one (Gardiner 1983).

### *6.3.3 Diffusion Models*

How is it possible to use the random walk model (6.28) as a diffusion model (i.e., a model for the transport of any inert substance due to the random motions of any fluid)? In particular, how can we calculate the evolution of the concentration of any inert substance? Let us address these questions in the following.

**Definition of Concentration.** First, we have to define the mean concentration $C$ of any substance considered. The general relation of the mean concentration of a substance to the statistics of random trajectories can be found elsewhere (Durbin 1983, Seinfeld & Pandis 2006). Here, we will focus on a simple but relevant case: we consider an instantaneous emission from a point source, i.e., the emission of a mass $M$ at time zero at a fixed position $y_0$. For this case, the mean concentration $C$ is given by $M$ times the PDF $f_n(y)$ for finding a parcel at a step $n$ at a position $y$,

$$C_n = M\, f_n(y). \tag{6.46}$$

By using the result (6.36) for the PDF of $y_n$ we find the mean concentration to be given by

$$C_n = \frac{M}{\sqrt{2\pi}\,\sigma_n} \exp\left\{ -\frac{(y - y_0)^2}{2\sigma_n{}^2} \right\}. \tag{6.47}$$

Here, the mean value is equal to the fixed position $y_0$, and the variance is given by $\sigma_n^2 = D\, n\, \Delta t$. Equation (6.47) describes the temporal evolution of the mean concentration in one dimension: the $y$-axis. It is worth noting that the total mass

involved remains constant for all $n$,

$$\int C_n \, dy = M \,, \tag{6.48}$$

as may be seen by using the normalization property of the normal PDF.

**Initial Condition.** Is Eq. (6.47) consistent with the assumed initial condition? The direct calculation of the initial concentration $C_0$ according to Eq. (6.47) is problematic because of $\sigma_0^2 = 0$. The best way to calculate the initial concentration is to consider the limit $\sigma_n^2 \to 0$. According to Eq. (6.47), the initial concentration at is then given by

$$C_0 = M \, \delta(y - y_0), \tag{6.49}$$

where we use the expression

$$\delta(y - y_0) = \lim_{\sigma_n \to 0} \frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left\{ -\frac{(y - y_0)^2}{2\sigma_n^{\,2}} \right\}. \tag{6.50}$$

This function is another representation of the delta function, see the discussion in Sect. 4.2.2 and exercise 6.3.3. This function is a normal PDF centered at $y_0$ with vanishing variance. This means that the only nonzero concentration is found at $y = y_0$. Due to the fact that the integral of $C_0$ over $y$ is equal to $M$, we find that all the mass $M$ is emitted at $y = y_0$, as assumed in the preceding paragraph.

**Boundary Effects.** The concentration formula (6.47) assumes that there are no boundaries, which is not the usual case if we consider diffusion processes. Let us consider two relatively simple examples to show how the effect of boundaries can be accounted for. A first case is given if there is a total reflection of material at $y = 0$. The presence of such a totally reflecting boundary can be taken into account by assuming that there is a hypothetical source at $y = -y_0$. The contributions of the sources at $y = y_0$ and $y = -y_0$ then result in the concentration

$$C_n = \frac{M}{\sqrt{2\pi}\,\sigma_n} \left\{ \exp\left\{ -\frac{(y - y_0)^2}{2\sigma_n^{\,2}} \right\} + \exp\left\{ -\frac{(y + y_0)^2}{2\sigma_n^{\,2}} \right\} \right\}. \tag{6.51}$$

To check the validity of this formula, let us integrate $C_n$ over the range $0 \le y < \infty$,

$$\begin{aligned} \int_0^\infty C_n \, dy &= \frac{M}{\sqrt{2\pi}\,\sigma_n} \left\{ \int_0^\infty \exp\left\{ -\frac{(y - y_0)^2}{2\sigma_n^{\,2}} \right\} dy + \int_0^\infty \exp\left\{ -\frac{(y + y_0)^2}{2\sigma_n^{\,2}} \right\} dy \right\} \\ &= \frac{M}{\sqrt{\pi}} \left\{ \int_{-y_0/\sqrt{2\sigma_n^{\,2}}}^{\infty} e^{-r^2} dr - \int_{-y_0/\sqrt{2\sigma_n^{\,2}}}^{-\infty} e^{-s^2} ds \right\} = \frac{M}{\sqrt{\pi}} \left\{ \int_{-\infty}^{-y_0/\sqrt{2\sigma_n^{\,2}}} e^{-s^2} ds + \int_{-y_0/\sqrt{2\sigma_n^{\,2}}}^{\infty} e^{-s^2} ds \right\} \\ &= \frac{M}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-s^2} ds = M. \end{aligned} \tag{6.52}$$

Here, we introduced the variables $r = (y - y_0) / (2^{1/2}\,\sigma_n)$ and $s = -(y + y_0) / (2^{1/2}\,\sigma_n)$. The last expression makes use of Eq. (4.70). Hence, all the material is conserved for this case of a totally reflecting boundary. As a second example, let us consider the case of a totally absorbing boundary at $y = 0$. The presence of such a boundary can be accounted for by assuming that there is a hypothetical source at $y = -y_0$. The difference to the approach applied to account for a totally reflecting boundary is that we have to consider now the difference of both distributions to ensure that $C_n = 0$ at the boundary $y = 0$,

$$C_n = \frac{M}{\sqrt{2\pi}\,\sigma_n}\left\{\exp\left\{-\frac{(y - y_0)^2}{2{\sigma_n}^2}\right\} - \exp\left\{-\frac{(y + y_0)^2}{2{\sigma_n}^2}\right\}\right\}. \tag{6.53}$$

To find out how much material is left in the accessible domain $0 \le y < \infty$ for this case we integrate again $C_n$, which leads to

$$\int_0^\infty C_n\,dy = \frac{M}{\sqrt{\pi}} \int_{-y_0/\sqrt{2{\sigma_n}^2}}^{y_0/\sqrt{2{\sigma_n}^2}} e^{-s^2}\,ds. \tag{6.54}$$

The proof of this formula is the concern of exercise 6.3.6. As expected, the total amount of material is not conserved in general for this case because the integral is smaller than $\pi^{1/2}$. It is worth noting that $C_n$ integrates to $M$ if $y_0 / (2^{1/2}\,\sigma_n) \to \infty$.

**Ground Concentrations.** The mean concentration development in time $t = n\,\Delta t$ can be used to find a corresponding two-dimensional concentration in a $x$-$y$ plane, where $x$ and $y$ refer to the horizontal and vertical coordinates. The latter can be achieved by assuming that the substance is transported along the $x$-direction with a constant velocity $U = x_n / (n\,\Delta t)$. By using the relation $n\,\Delta t = x_n / U$, the variance ${\sigma_n}^2 = D\,n\,\Delta t = D\,x_n / U$ becomes a function of $x_n$. Let us calculate the ground concentration at $y = 0$ for the case without boundary to illustrate the use of this approach. The results to be obtained can be applied to the case of a totally reflecting boundary by multiplying the ground concentration for the case without boundary with two. The consideration of the ground concentration for the totally absorbing boundary case does not make sense because this ground concentration is equal to zero. According to Eq. (6.47) and ${\sigma_n}^2 = D\,x_n / U$, the ground concentration for the case without boundary is given by

$$C_n = M\sqrt{\frac{U}{2\pi D\,x_n}}\exp\left\{-\frac{1}{2}\frac{U\,{y_0}^2}{D\,x_n}\right\}. \tag{6.55}$$

It is convenient to introduce the nondimensional positions $x_{*n} = D\,x_n / (U\,{y_0}^2)$ and concentrations $C_{*n} = C_n y_0 / M$. Then, Eq. (6.55) can be written as

$$C_{*n} = \frac{1}{\sqrt{2\pi\,x_{*n}}}\exp\left\{-\frac{1}{2\,x_{*n}}\right\}. \tag{6.56}$$

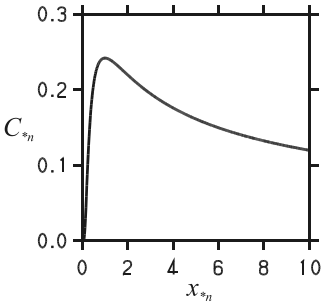



**Fig. 6.6.** The normalized ground concentration $C_{*_n}$ along the normalized horizontal coordinate $x_{*_n}$ for the case without boundary (see Eq. (6.56)).

An illustration of the normalized ground concentration distribution is given in Fig. 6.6. It is reasonable that the $C_{*_n}$ curve shows a maximum. The plume does hardly hit the ground for very small distances $x_{*_n}$, and the concentration $C_{*_n}$ has to become smaller for growing large distances $x_{*_n}$ because of the ongoing substance diffusion. The maximum position can be calculated by considering the first two derivatives of Eq. (6.56). We find for this case

$$\frac{dC_{*_n}}{dx_{*_n}} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2x_{*_n}}\right\}\left(\frac{1}{2x_{*_n}^{1/2}x_{*_n}^{2}} - \frac{1}{2x_{*_n}^{3/2}}\right) = \frac{1-x_{*_n}}{2x_{*_n}^{2}} C_{*_n},$$

$$\frac{d^2C_{*_n}}{dx_{*_n}^{2}} = \frac{(1-x_{*_n})^2}{4x_{*_n}^{4}} C_{*_n} + C_{*_n}\left(-\frac{1}{x_{*_n}^{3}} + \frac{1}{2x_{*_n}^{2}}\right) = C_{*_n}\left[\frac{(1-x_{*_n})^2}{4x_{*_n}^{4}} + \frac{x_{*_n}-2}{2x_{*_n}^{3}}\right]. \tag{6.57}$$

The first-order derivative disappears for $x_{*_n} = 1$, and the second-order derivative is negative at $x_{*_n} = 1$. Consequently, the concentration has a maximum at $x_{*_n} = 1$. The maximum value of $C_{*_n}$ is given by

$$C_{*_n} = \frac{e^{-1/2}}{\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi e}} = 0.242. \tag{6.58}$$

Such analytical results give a valuable guideline for the evaluation of maximum ground concentrations depending on the nature of boundary.

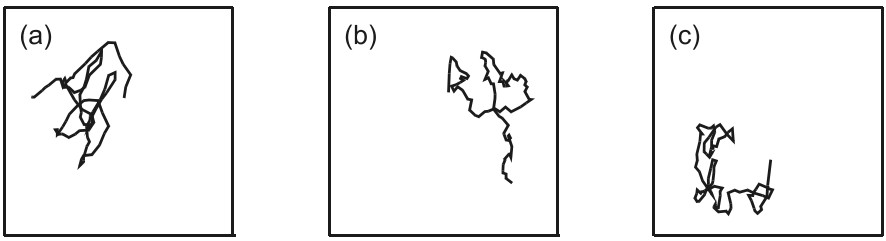



**Fig. 6.7.** Three examples for two-dimensional Brownian motion of particles.

## 6.4 Brownian Motion

Next, let us consider the modeling of Brownian motion. Brownian motion of particles refers to the following phenomenon: fine particles, when suspended in water, move in an irregular way (see, e.g., the illustration in Fig. 6.7). In 1827, the botanist Robert Brown investigated such motions by considering the suspension of small pollen grains in water (therefore, all such irregular particle motions are called Brownian motion). Brown was interested to understand whether the chaotic motion of pollen grains is a manifestation of life. By repeating the experiment with particles of dust, he was able to rule out that the motion was due to pollen particles being alive, although the origin of the motion was yet to be explained. The relevance of this problem arises from the fact that the modeling of Brownian motion provides the basis for the development of models for molecular and fluid flow motion. However, the analysis of such models requires knowledge about the treatment of joint PDFs and joint PDF transport equations and their consequences. Therefore, such molecular and fluid motion models will be considered in Chap. 10 after introducing the required mathematical concepts. Here, we will consider this problem because of two reasons. First, this discussion follows the approach used in Chap. 5 to continue after the discussion of linear first-order equations with the discussion of second-order difference equations. The second reason for considering this problem is that the Brownian motion model represents an extension of the diffusion model discussed in Sect. 6.3.

### *6.4.1 Brownian Motion Model*

**Brownian Motion Model.** The explanation for the Brownian motion problem was published by Einstein (1905). Einstein explained that the irregular motion of pollen grains is caused by the exceedingly frequent impacts on the pollen grain of the incessantly moving molecules of liquid in which it is suspended. Einstein developed a mathematical model (a Fokker-Planck equation: see the discussion in Chap. 10) for the explanation of Brownian motions. A few years after Einstein's explanation, Langevin (1908) presented the corresponding stochastic equations for the Brownian motion of particles. We will follow here Langevin's approach by considering the following stochastic difference equation system

$$\frac{x_n - x_{n-1}}{\Delta t} = v_{n-1}, \tag{6.59a}$$

$$\frac{v_n - v_{n-1}}{\Delta t} = -\frac{1}{\tau}\left(v_{n-1} - \sqrt{D}\,\frac{\Delta W_{n-1}}{\Delta t}\right). \tag{6.59b}$$

Here, $x_n$ and $v_n$ refer to the position and velocity of a Brownian particle, respectively. No assumption is made regarding the time interval $\Delta t$. $D$ denotes again a diffusion coefficient, $\tau$ represents a characteristic time scale, and the change of the Wiener process is defined by $\Delta W_{n-1} = (\Delta t)^{1/2}\, \varepsilon_{n-1}$. The last term on the right-hand side of Eq. (6.59b) provides a random input (as a model for the random impacts of water molecules on a pollen grain), and the first term on the right-hand side of Eq. (6.59b) models the relaxation of the pollen velocity due to the damping influence of surrounding water molecules: for a positive (negative) velocity $v_{n-1}$ there will be a decrease (increase) of the velocity due to this relaxation term.

**Linear Second-Order Difference Equation.** Let us show first that the model (6.59) extends the discussion of a linear first-order difference equation in Sect. 6.3 by the consideration of a linear second-order difference equation. Equation (6.59b) can be also written as

$$v_n = \left(1 - \frac{\Delta t}{\tau}\right) v_{n-1} + \frac{1}{\tau}\sqrt{D}\,\Delta W_{n-1} = \left(1 - \frac{\Delta t}{\tau}\right)\frac{x_n - x_{n-1}}{\Delta t} + \frac{1}{\tau}\sqrt{D}\,\Delta W_{n-1}, \tag{6.60}$$

where Eq. (6.59a) was used to replace $v_{n-1}$ on the right-hand side. The use of this equation at $n - 1$ on the right-hand side of Eq. (6.59a) results in

$$\frac{x_n - x_{n-1}}{\Delta t} = \left(1 - \frac{\Delta t}{\tau}\right)\frac{x_{n-1} - x_{n-2}}{\Delta t} + \frac{1}{\tau}\sqrt{D}\,\Delta W_{n-2}. \tag{6.61}$$

This equation represents a linear second-order difference equation for $x_n$,

$$x_n = x_{n-1} + \left(1 - \frac{\Delta t}{\tau}\right)(x_{n-1} - x_{n-2}) + \frac{\Delta t}{\tau}\sqrt{D\,\Delta t}\;\varepsilon_{n-2}. \tag{6.62}$$

Here, the noise term is written by making use of $\Delta W_{n-2} = (\Delta t)^{1/2}\, \varepsilon_{n-2}$.

**Comparison with Diffusion Model.** Let us consider the difference between Eq. (6.62) and the diffusion model (6.28). We can write Eq. (6.28) in terms of $x_n$

$$x_n = x_{n-1} + \sqrt{D\,\Delta t}\;\varepsilon_{n-1}, \tag{6.63}$$

where $r = (D\,\Delta t)^{1/2}$ is used. The random variable $\varepsilon_{n-2}$ can be replaced by $\varepsilon_{n-1}$, which does not make a difference. It may be seen that the setting $\Delta t = \tau$ reduces the Brownian motion model (6.62) to the diffusion model (6.63), this means the diffusion model represents a coarse Brownian motion model in which the time step $\Delta t$ is equal to the typical relaxation time $\tau$. The relevant difference between both models is given by the appearance of the second term on the right-hand side of Eq. (6.62). What is the advantage of involving this term? The consideration of this term is equivalent to the consideration of the first term on the right-hand side of Eq. (6.60). Without involving this term, the velocity $v_n$ is proportional to $\Delta W_{n-1}/\Delta t$. It was shown in Sect. 6.3.2 that the variance of $\Delta W_{n-1}/\Delta t$ does not exist

for a continuous diffusion process (i.e., for $\Delta t \to 0$). Apart from that, the diffusion velocities would be uncorrelated due to the properties of $\Delta W_n$. Thus, the neglect of the first term on the right-hand side of Eq. (6.60) does not result in a sound model for diffusion in reality. In contrast to that, the consideration of this first term enables the modeling of correlated velocities that have a finite variance.

### *6.4.2 Discrete Brownian Motion Statistics*

**Joint PDF.** Let us consider the consequences of the Brownian motion model (6.59). By adopting the same reasoning as with regard to $y_n$ in Sect. 6.2, it is possible to show that both $x_n$ and $v_n$ are normally distributed if $x_0$ and $v_0$ are normally distributed as assumed here. However, this does not mean that the joint process ($x_n$, $v_n$) is also normally distributed. To clarify this question, we have to ask whether all the consequences (10.43) of a bivariate normal PDF for central moments of third-order, fourth-order, fifth-order, and sixth-order are satisfied (see the corresponding discussion in Chap. 10). By solving Eq. (6.59) for a number of $10^7$ realizations, one finds that the magnitude of the deviations between the calculated and theoretical central moments of third-order, fourth-order, fifth-order, and sixth-order is smaller than 0.0085 for all the 22 conditions considered. The relative error can be calculated for the nonzero fourth-order and sixth-order moments. This calculation shows that the magnitude of the relative error is smaller than 0.22%. Therefore, we find in this way evidence that the joint process ($x_n$, $v_n$) is indeed normally distributed.

**Solutions.** The best way to find the means and variances of the joint normal PDF of $x_n$ and $v_n$ is to calculate the solutions of Eq. (6.59). For doing this we use the abbreviations $a = 1 - \Delta t / \tau$ and $r_B = (D\,\Delta t / \tau^2)^{1/2}$, where $r_B$ is the noise intensity in the Brownian motion velocity equation. Equations (6.59) read then

$$x_n = x_{n-1} + \Delta t\, v_{n-1}, \tag{6.64a}$$

$$v_n = a\, v_{n-1} + r_B\, \varepsilon_{n-1}. \tag{6.64b}$$

For the first three $v_n$ we find according to Eq. (6.64b)

$$v_1 = a\, v_0 + r_B\, \varepsilon_0, \tag{6.65a}$$

$$v_2 = a\, v_1 + r_B\, \varepsilon_1 = a^2 v_0 + r_B \left(a\, \varepsilon_0 + \varepsilon_1\right), \tag{6.65b}$$

$$v_3 = a\, v_2 + r_B\, \varepsilon_2 = a^3 v_0 + r_B \left(a^2\, \varepsilon_0 + a\, \varepsilon_1 + \varepsilon_2\right). \tag{6.65c}$$

Correspondingly, the solution of Eq. (6.64b) is given by the expression

$$v_n = a^n\, v_0 + r_B \left(a^{n-1}\, \varepsilon_0 + \cdots + a^1\, \varepsilon_{n-2} + a^0\, \varepsilon_{n-1}\right). \tag{6.66}$$

The solution for $x_n$ can be similarly obtained by considering the implications of Eq. (6.64a) for $n = 1$, 2, and 3,

$$x_1 = x_0 + \Delta t\, v_0, \tag{6.67a}$$

$$x_2 = x_1 + \Delta t\, v_1 = x_0 + \Delta t\,(v_0 + v_1), \tag{6.67b}$$

$$x_3 = x_2 + \Delta t\, v_2 = x_0 + \Delta t\,(v_0 + v_1 + v_2). \tag{6.67c}$$

By following this pattern, the solution $x_n$ is found to be given by

$$\begin{aligned} x_n &= x_0 + \Delta t\left(v_0 + v_1 + v_2 + \cdots + v_{n-1}\right) \\ &= x_0 + \Delta t\Big(v_0\,(1 + a + a^2 + \cdots a^{n-1}) + r_B\Big[\varepsilon_0\,(1 + a + a^2 + \cdots a^{n-2}) \\ &\qquad + \varepsilon_1\,(1 + a + a^2 + \cdots a^{n-3}) + \cdots + \varepsilon_{n-3}\,(1 + a) + \varepsilon_{n-2}\Big]\Big) \\ &= x_0 + v_0 \Delta t\,\frac{1 - a^n}{1 - a} + r_B \Delta t\left[\varepsilon_0\,\frac{1 - a^{n-1}}{1 - a} + \varepsilon_1\,\frac{1 - a^{n-2}}{1 - a} + \cdots + \varepsilon_{n-3}\,\frac{1 - a^2}{1 - a} + \varepsilon_{n-2}\,\frac{1 - a}{1 - a}\right] \\ &= x_0 + \tau\, v_0\,(1 - a^n) + \\ &\quad + \sqrt{D\,\Delta t}\left[\varepsilon_0 (1 - a^{n-1}) + \varepsilon_1 (1 - a^{n-2}) + \cdots + \varepsilon_{n-3}(1 - a^2) + \varepsilon_{n-2}(1 - a)\right]. \end{aligned} \tag{6.68}$$

The second line results from the sum of Eqs. (6.65). The third line applies the identity (6.10), and the last expression applies the relations

$$\frac{\Delta t}{1 - a} = \frac{\Delta t}{1 - \left(1 - \Delta t / \tau\right)} = \tau, \qquad \frac{r_B\, \Delta t}{1 - a} = \frac{\sqrt{D\,\Delta t / \tau^2}\ \Delta t}{1 - \left(1 - \Delta t / \tau\right)} = \sqrt{D\,\Delta t}\,. \tag{6.69}$$

**Means and Variances.** The means of $x_n$ and $v_n$ follow from taking the means of Eqs. (6.66) and (6.68), respectively,

$$\left\langle x_n \right\rangle = \left\langle x_0 \right\rangle + \tau \left\langle v_0 \right\rangle (1 - a^n), \tag{6.70a}$$

$$\left\langle v_n \right\rangle = a^n \left\langle v_0 \right\rangle. \tag{6.70b}$$

The fluctuations of $x_n$ and $v_n$ are then found by taking the differences between Eqs. (6.68) and (6.70a), and (6.66) and (6.70b),

$$\begin{aligned} \widetilde{x}_n &= \widetilde{x}_0 + \tau\, \widetilde{v}_0\,(1 - a^n) + \\ &\quad + \sqrt{D\,\Delta t}\left[\varepsilon_0 (1 - a^{n-1}) + \varepsilon_1 (1 - a^{n-2}) + \cdots + \varepsilon_{n-3}(1 - a^2) + \varepsilon_{n-2}(1 - a)\right], \end{aligned} \tag{6.71a}$$

$$\widetilde{v}_n = a^n\, \widetilde{v}_0 + \sqrt{\frac{D\,\Delta t}{\tau^2}}\left(a^{n-1}\, \varepsilon_0 + \cdots + a^1\, \varepsilon_{n-2} + a^0\, \varepsilon_{n-1}\right). \tag{6.71b}$$

We assume that $\widetilde{x}_0$ is independent of $\widetilde{v}_0$, and $\widetilde{x}_0$ and $\widetilde{v}_0$ are independent of all $\varepsilon_k$. The variances of $x_n$ and $v_n$ can be obtained by multiplication and averaging of the corresponding fluctuations $\widetilde{x}_0$ and $\widetilde{v}_0$. As shown in exercises 6.4.1 and 6.4.2, we

obtain in this way the expressions

$$\left\langle \tilde{x}_n^{\,2} \right\rangle = \left\langle \tilde{x}_0^{\,2} \right\rangle + \tau^2 \left\langle \tilde{v}_0^{\,2} \right\rangle (1-a^n)^2 + D\left[ n\,\Delta t - 2\,\tau\,(1-a^n) + \tau \frac{1-a^{2n}}{1+a} \right], \tag{6.72a}$$

$$\left\langle \tilde{x}_n \tilde{v}_n \right\rangle = \tau \left\langle \tilde{v}_0^{\,2} \right\rangle (1-a^n)\, a^n + D\left[ 1-a^n - \frac{1-a^{2n}}{1+a} \right], \tag{6.72b}$$

$$\left\langle \tilde{v}_n^{\,2} \right\rangle = a^{2n} \left\langle \tilde{v}_0^{\,2} \right\rangle + \frac{D}{\tau} \frac{1-a^{2n}}{1+a}. \tag{6.72c}$$

As shown in exercise 6.4.3, these relations contain the statistics of the diffusion model (6.63) as a specific case.

### *6.4.3 Continuous Brownian Motion Statistics*

**Continuous Statistics.** A relevant question, which was already considered in Sect. 6.3.1 regarding the position variance, is whether the Brownian motion model statistics represent a consistent continuous time model, or in other words whether $n$ and the time interval $\Delta t$ do only appear for $\Delta t / \tau \to 0$ in the combination $t = n\,\Delta t$ (the appearance of terms like $n\,(\Delta t)^2$ would indicate an inappropriate model formulation). First, we consider

$$\lim_{\Delta t \to 0} a^n = \lim_{\Delta t \to 0} \left( 1 - \frac{\Delta t}{\tau} \right)^n = \lim_{\Delta t \to 0} \exp\left\{ n \ln\left( 1 - \frac{\Delta t}{\tau} \right) \right\} = \exp\left\{ -\frac{n\,\Delta t}{\tau} \right\} = e^{-t/\tau}. \tag{6.73}$$

Here, the logarithmic function was replaced by the first term of its Taylor series,

$$\lim_{\Delta t \to 0} \ln\left( 1 - \frac{\Delta t}{\tau} \right) = -\frac{\Delta t}{\tau}. \tag{6.74}$$

In terms of this limit for $a^n$ we obtain the means (6.70) in the limit $\Delta t / \tau \to 0$ as

$$\left\langle x \right\rangle = \left\langle x_0 \right\rangle + \tau \left\langle v_0 \right\rangle (1 - e^{-t/\tau}), \tag{6.75a}$$

$$\left\langle v \right\rangle = e^{-t/\tau} \left\langle v_0 \right\rangle. \tag{6.75b}$$

Here, $x_n$ and $v_n$ are replaced by $x$ and $v$, respectively, because we consider now functions of $t$. Therefore, these expressions represent indeed functions of $t$ in the limit $\Delta t / \tau \to 0$. The consistency with the initial values can be seen by setting $t = 0$. The means (6.75) are related in the continuous time limit by the relations

$$\left\langle v \right\rangle = \frac{d\left\langle x \right\rangle}{dt}. \tag{6.76}$$

The variances (6.72), too, become consistent functions of $t$ in the limit $\Delta t / \tau \to 0$. In terms of Eq. (6.73) we find

$$\left\langle \tilde{x}^2 \right\rangle = \left\langle \tilde{x}_0^{\,2} \right\rangle + \tau^2 \left\langle \tilde{v}_0^{\,2} \right\rangle (1 - e^{-t/\tau})^2 + D\tau \left[ \frac{t}{\tau} - 2(1 - e^{-t/\tau}) + \frac{1 - e^{-2t/\tau}}{2} \right], \tag{6.77a}$$

$$\left\langle \tilde{x}\,\tilde{v} \right\rangle = \tau \left\langle \tilde{v}_0^{\,2} \right\rangle (1 - e^{-t/\tau})\, e^{-t/\tau} + D \left[ 1 - e^{-t/\tau} - \frac{1 - e^{-2t/\tau}}{2} \right], \tag{6.77b}$$

$$\left\langle \tilde{v}^2 \right\rangle = e^{-2t/\tau} \left\langle \tilde{v}_0^{\,2} \right\rangle + \frac{D}{2\tau} (1 - e^{-2t/\tau}). \tag{6.77c}$$

We note that the term $1 + a$ has to be replaced by 2. As done regarding the means, $x_n$ and $v_n$ are replaced here by $x$ and $v$, respectively. The variances (6.77) obtained provide the correct initial values, as may be seen by setting $t = 0$. We note that we have the following relation in the continuous time limit,

$$\left\langle \tilde{x}\,\tilde{v} \right\rangle = \frac{1}{2} \frac{d \left\langle \tilde{x}^2 \right\rangle}{dt}. \tag{6.78}$$

Hence, the covariance is controlled by the variance of $x(t)$, such that there is no reason to consider the covariance regarding the following comparisons.

**Comparison with Diffusion Model: Asymptotic Limit.** Let us compare the asymptotic features of the Brownian motion model and the diffusion model. In the limit $\Delta t / \tau \to \infty$, the position statistics that are implied by the Brownian motion model are given by

$$\left\langle x \right\rangle = \left\langle x_0 \right\rangle + \tau \left\langle v_0 \right\rangle, \tag{6.79a}$$

$$\left\langle \tilde{x}^2 \right\rangle = Dt. \tag{6.79b}$$

These results agree with the implications of the diffusion model (6.63), except that Eq. (6.63) applies a zero mean initial velocity because of $x_n - x_{n-1} = D^{1/2}\, \Delta W_{n-1}$. For the velocity statistics the Brownian motion model provide for $\Delta t / \tau \to \infty$

$$\left\langle v \right\rangle = 0. \tag{6.80a}$$

$$\left\langle \tilde{v}^2 \right\rangle = \frac{D}{2\tau}. \tag{6.80b}$$

Equation (6.80a) agrees with the consequences of the diffusion model. The availability of Eq. (6.80b) represents a significant difference to the Brownian motion model, which does not offer such a relation (see the discussion on the properties of the Wiener process in Sect. 6.3.3). The advantage of Eq. (6.80b) is that we can use this so-called Einstein relation for the calculation of the diffusion coefficient $D$ provided the characteristic relaxation time scale $\tau$ and the variance are known.

For $\tau$, Einstein used Stokes' Law (see the discussion in Sect. 3.3.3) given by

$$\tau = \frac{m}{6\pi\mu r}. \tag{6.81}$$

Here, $r$ denotes the radius of a spherical particle, $m$ is the particle mass, and $\mu$ is the dynamic viscosity. The variance in Eq. (6.80b) is determined according to the Equipartition Law in one dimension,

$$\frac{m}{2}\left\langle \tilde{v}^2 \right\rangle = \frac{k}{2}T. \tag{6.82}$$

Here, $k$ represents Boltzmann's constant, and $T$ refers to the absolute temperature. The combination of the last two relations with Eq. (6.80b) does allow us then to calculate the diffusion coefficient,

$$D = \frac{kT}{3\pi\mu r}. \tag{6.83}$$

Therefore, the diffusion coefficient depends linearly on the temperature $T$ if the models (6.81) and (6.82) are applied.

**Comparison with Diffusion Model: Transition to Asymptotic Limit.** Next, let us compare the transition to the asymptotic limit according to the Brownian motion model and diffusion model. To apply the Brownian motion model we have to specify the initial velocity statistics $<v_0>$ and $<\tilde{v}_0^{\,2}>$. For simplicity, we assume that the initial velocity statistics are given by the equilibrium values

$$\left\langle v_0 \right\rangle = 0, \qquad \left\langle \tilde{v}_0^{\,2} \right\rangle = \frac{D}{2\tau}. \tag{6.84}$$

Under these conditions, Eqs. (6.75b) and (6.77c) show that the mean and variance are also given by their equilibrium values, $<v> = 0$ and $<\tilde{v}^2> = D/(2\,\tau)$. According to Eqs. (6.75a) and (6.77a), the mean and variances of $x$ are then given by

$$\left\langle x \right\rangle = \left\langle x_0 \right\rangle, \tag{6.85a}$$

$$\begin{aligned}\left\langle \tilde{x}^2 \right\rangle &= \left\langle \tilde{x}_0^{\,2} \right\rangle + \frac{D\tau}{2}(1-e^{-t/\tau})^2 + D\tau\left[\frac{t}{\tau} - 2(1-e^{-t/\tau}) + \frac{1-e^{-2t/\tau}}{2}\right] \\ &= \left\langle \tilde{x}_0^{\,2} \right\rangle + Dt + D\tau(1-e^{-t/\tau})\left[\frac{1-e^{-t/\tau}}{2} - 2 + \frac{1+e^{-t/\tau}}{2}\right] \\ &= \left\langle \tilde{x}_0^{\,2} \right\rangle + Dt - D\tau(1-e^{-t/\tau}).\end{aligned} \tag{6.85b}$$

To illustrate the behavior of the position variance, we introduce the normalized deviation from the initial variance

$$B_{xx} = \frac{\left\langle \tilde{x}^2 \right\rangle - \left\langle \tilde{x}_0^{\,2} \right\rangle}{D\tau}. \tag{6.86}$$

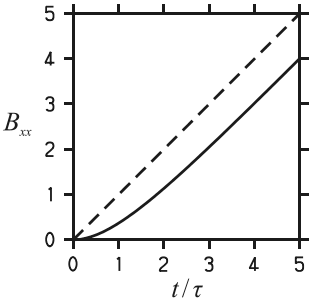



**Fig. 6.8.** The *solid line* shows $B_{xx}$ as function of $t/\tau$ according to the Brownian motion formula (6.87). The *dashed line* shows $B_{xx}^{DM}$ according to Eq. (6.88) for the diffusion model.

This quantity is given according to Eq. (6.85b) by

$$B_{xx} = \frac{t}{\tau} + e^{-t/\tau} - 1. \tag{6.87}$$

The corresponding behavior $B_{xx}^{DM}$ according to the diffusion model, which is denoted by the superscript *DM*, is given by

$$B_{xx}^{DM} = \frac{t}{\tau}. \tag{6.88}$$

Figure 6.8 shows $B_{xx}$ and $B_{xx}^{DM}$ as functions of $t/\tau$. The behavior of the Brownian motion model is the correct behavior, which is in agreement with measurements (Durbin 1983). The coarse diffusion model, which applies large time steps $\Delta t = \tau$, does correctly describe the asymptotic behavior $t/\tau \to \infty$, but the prediction of the diffusion model is incorrect for relatively small $t/\tau$, which is the most important period where most of the changes occur. In addition to the advantage described in the preceding paragraph, this comparison reveals another significant advantage of the Brownian motion model compared to the diffusion model.

## 6.5 Population Dynamics

Next, let us consider the stochastic modeling of the evolution of populations. In this way, we follow the organization of Chap. 5 by considering nonlinear models after the discussion of linear models. In particular, we will consider the extension of the logistic population model discussed in Chap. 5 to a stochastic population model. One purpose of the following discussion is to illustrate the scenarios that can be described if randomness is involved in nonlinear equations. Another purpose is the illustration of typical mathematical problems that appear in analyses of nonlinear stochastic equations. The analyses of many other problems lead to very similar questions, as given, for example, with regard to the application of nonlinear stochastic equations for turbulence.

### 6.5.1 A Stochastic Logistic Model

**Logistic Model.** The model considered in the following is the discrete logistic model (5.97), which we write in terms of $P_n$ instead of $y_n$ used in Sect. 5.4.2,

$$P_n = P_{n-1} + a P_{n-1}\left(1 - \frac{P_{n-1}}{K}\right). \tag{6.89}$$

There are two equilibrium states $P_n = 0$ and $P_n = K$ that can be realized depending on the initial population $P_0$: these equilibrium states imply $P_n - P_{n-1} = 0$, such that $P_n$ will not change anymore. $K$ is the carrying capacity, and the model parameter $a$ determines the transition rate to the equilibrium state. To refer to the meaning of $a$ we replace it by $a = \Delta t / T$, where $\Delta t$ denotes a time interval (which does not have to be small), and $T$ is a characteristic time scale. By using the definition $a = \Delta t / T$, we can write Eq. (6.89) as

$$\frac{P_n - P_{n-1}}{\Delta t} = \frac{P_{n-1}}{T}\left(1 - \frac{P_{n-1}}{K}\right). \tag{6.90}$$

**Stochastic Logistic Model.** The evolution of populations is often affected by fluctuations that may be caused by variations of the available food, the impact of natural enemies, diseases, or weather conditions. Here, we have two possibilities: we can randomize $K$ or $T$. The randomization of $K$ is questionable from a conceptual point of view because $K$ is seen to represent an upper limit for the population density that can be supported with food over a long term in a given area: see the discussions in Sects. 7.4 and 7.5. The randomization of $K$ is also questionable regarding the fact that a random model for $K$ may result in a negative population density $K$, which does not make sense. To simplify the following discussion we assume that $K = 1$. The influence of $K$ can be covered by dividing the population equation (6.90) by $K$ and introducing a new variable $P_n / K$, so that the population equation becomes independent of $K$. The consequence of setting $K = 1$ is that the population is now bounded, $0 \le P_n \le 1$. To account for the effect of randomness on the population dynamics we randomize $T$ here. The consideration of negative and positive values of the growth time $T$ represents an appropriate mean to reflect varying conditions for a population development. However, the direct randomization of $T$ does represent a very questionable approach because of the potential problem to divide by a zero time scale value $T$. Thus, we will assume that $T^{-1}$ is normally distributed,

$$T^{-1} = \mu + \sigma \frac{\Delta W_{n-1}}{\Delta t}. \tag{6.91}$$

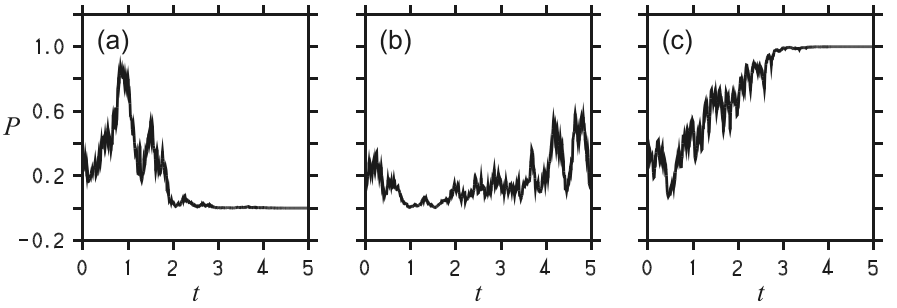



**Fig. 6.9.** Three realizations of the stochastic population model (6.92), where $t$ refers to time in years. The simulation conditions are described in Sect. 6.5.2.

We apply $\Delta W_{n-1} = (\Delta t)^{1/2}\, \varepsilon_{n-1}$ as before, and $\mu$ and $\sigma$ denote model parameters that will be specified below. The combination of the population model (6.90) with this model for $T^{-1}$ and $K = 1$ with leads then to the model

$$\frac{P_n - P_{n-1}}{\Delta t} = P_{n-1}\left(1 - P_{n-1}\right)\left(\mu + \sigma \frac{\Delta W_{n-1}}{\Delta t}\right). \tag{6.92}$$

The stochastic logistic model (6.92) does not represent the only stochastic version of the logistic model. Different other stochastic models can be designed by using certain birth and death rate models in equations for the population density PDF (see Allen 2003). The structure of Eq. (6.92) corresponds to the structure of the diffusion model (6.28) because the derivative of the variable considered is proportional to $\Delta W_{n-1} / \Delta t$. The suitability of such an assumption is illustrated in terms of exercise 6.3.2.

**Realizations.** Three realizations of the stochastic population model (6.92) are shown in Fig. (6.9). The realizations considered illustrate an interesting property of the population model. Without randomizing $T^{-1}$, all trajectories with $0 < P_0 < 1$ would be finally attracted by the equilibrium state $P_n = 1$. However, by accounting for randomness, it is possible that the equilibrium state $P_n = 0$ will be realized, too: see Fig. 6.9a. There is no overshooting of the two equilibrium values as long as $\sigma$ is not too large.

### *6.5.2 One-Point Statistics and Correlations*

To understand the characteristic model features let us consider the development of one-point statistics (the population density PDF, mean, and variance) in time and the correlations that are implied by the stochastic model (6.92).

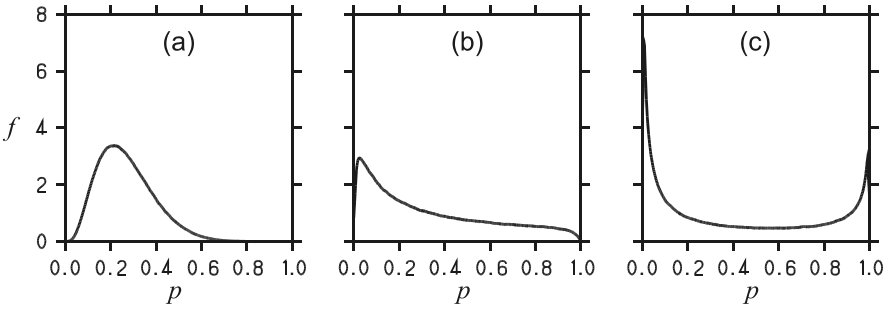



**Fig. 6.10.** The development of the population density PDF $f(p)$ in time $t$ according to the stochastic population model (6.92). The PDF is shown at $t = 0.1$, $t = 0.5$, and $t = 1$ in (a), (b), and (c), respectively.

**PDF.** The only way to investigate the nonlinear stochastic model statistics is to calculate the model solution numerically. This was done by using the parameter values $P_0 = 0.25$, $\mu = 1$, and $\sigma = 2$. The time step $\Delta t = 0.01$ was applied, and a number $N = 10^6$ of realizations was used. The PDF $f(p)$ was calculated at positions with a distance of 0.002 by using a filter size of 0.02. Here, $p$ refers to the sample space population density. The evolution of the population density PDF in time $t = n\,\Delta t$ is illustrated in Fig. 6.10. At $t = 0$, there is a delta function peak at $p = 0.25$. After a relatively short time $t = 0.1$, this delta peak is significantly distributed. The population probability flows then toward the equilibrium values $p = 0$ and $p = 1$. At $p = 0.5$, there is a nonzero probability of population values over the entire domain. The population probability flows toward $p = 0$, but there is no way to go beyond this value. This leads to the maximum of the population PDF $f(p)$ close to $p = 0$. For later times, the continuing flow of the population probability toward $p = 0$ and $p = 1$ leads to the development of sharp maximum values close to $p = 0$ and $p = 1$. Asymptotically, the PDF $f(p)$ develops delta-peak spikes close to $p = 0$ and $p = 1$, and the probability for other population values vanishes.

**Moments.** Next, let us analyze the mean and variance. In preparation of the derivation of an equation for the mean value of $P_n$, we write Eq. (6.92) as

$$P_n = P_{n-1} + P_{n-1}\left(1 - P_{n-1}\right)\left(\mu\,\Delta t + \sigma\,(\Delta t)^{1/2}\,\varepsilon_{n-1}\right), \tag{6.93}$$

where $\Delta W_{n-1} = (\Delta t)^{1/2}\,\varepsilon_{n-1}$ is applied. By averaging this relation, we obtain

$$\left\langle P_n \right\rangle = \left\langle P_{n-1} \right\rangle + \mu\,\Delta t\left\langle P_{n-1}\left(1 - P_{n-1}\right)\right\rangle = \left(1 + \mu\,\Delta t\right)\left\langle P_{n-1}\right\rangle - \mu\,\Delta t\left\langle {P_{n-1}}^2\right\rangle. \tag{6.94}$$

The term involving $\varepsilon_{n-1}$ does not provide a contribution to this equation because $\varepsilon_{n-1}$ is independent of $P_{n-1}$ ($P_{n-1}$ does only involve terms like $\varepsilon_{n-2}$, $\varepsilon_{n-3}$, ...). Therefore, we have $\langle P_{n-1}\,(P_{n-1} - 1)\,\varepsilon_{n-1}\rangle = \langle P_{n-1}\,(P_{n-1} - 1)\rangle \langle \varepsilon_{n-1}\rangle = 0$. Equation (6.94) for $\langle P_n\rangle$ is unclosed due to the appearance of the second-order moment $\langle {P_{n-1}}^2\rangle$.

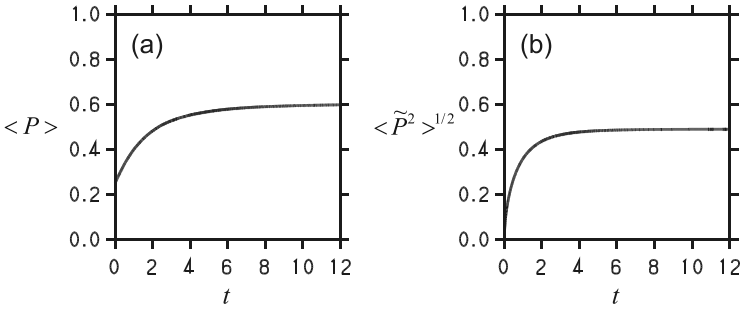



**Fig. 6.11.** The mean population density and standard deviation determined by the stochastic model (6.92) are shown in dependence on time $t$ (in years).

Let us derive an equation for the calculation of the second-order moment $<P_n^2>$. By squaring Eq. (6.93) we obtain

$$\begin{aligned} P_n^{\,2} &= P_{n-1}^{\,2} + 2P_{n-1}^{\,2}(1-P_{n-1})\left(\mu\Delta t + \sigma(\Delta t)^{1/2}\,\varepsilon_{n-1}\right) \\ &\quad + P_{n-1}^{\,2}(1-P_{n-1})^2\left(\mu^2(\Delta t)^2 + 2\mu\sigma(\Delta t)^{3/2}\,\varepsilon_{n-1} + \sigma^2\,\Delta t\,\varepsilon_{n-1}^{\,2}\right). \end{aligned} \tag{6.95}$$

After averaging this relation we find that the second-order moment of $P_n$ is determined by the equation

$$\begin{aligned} \left\langle P_n^{\,2}\right\rangle &= \left\langle P_{n-1}^{\,2}\right\rangle + 2\mu\Delta t\left\langle P_{n-1}^{\,2}(1-P_{n-1})\right\rangle + \mu^2(\Delta t)^2\left\langle P_{n-1}^{\,2}(1-P_{n-1})^2\right\rangle \\ &\quad + \sigma^2\,\Delta t\left\langle P_{n-1}^{\,2}(1-P_{n-1})^2\right\rangle. \end{aligned} \tag{6.96}$$

This expression was obtained by neglecting terms that involve $\varepsilon_{n-1}$, but the term involving $\varepsilon_{n-1}^{\,2}$ provides the nonzero last contribution. Regarding the limit $\Delta t \to 0$ we observe that the third term on the right-hand side can be neglected if $\Delta t \to 0$. Equation (6.96) includes unknowns in terms of the third-order and fourth-order moments involved. Equations for the third-order and fourth-order moments do again involve higher-order moments. Hence, a closed system for the calculation of the mean and the variance of $P_n$ cannot be obtained. The development of the mean population density and standard deviation in time $t = n\,\Delta t$ is shown in Fig. 6.11, where $N = 10^5$ realizations are applied. Both the mean and the standard deviation become constant asymptotically. The curve of the mean indicates that about 60% of the realizations do finally realize the value $P_n = 1$.

**Correlations.** The normalized correlation function of population densities at $t = n\,\Delta t$ and $s = m\,\Delta t$ is defined by

$$C(t,s) = \frac{\left\langle \widetilde{P}(t)\widetilde{P}(t+s)\right\rangle}{\left\langle \widetilde{P}(t)\widetilde{P}(t)\right\rangle}, \tag{6.97}$$

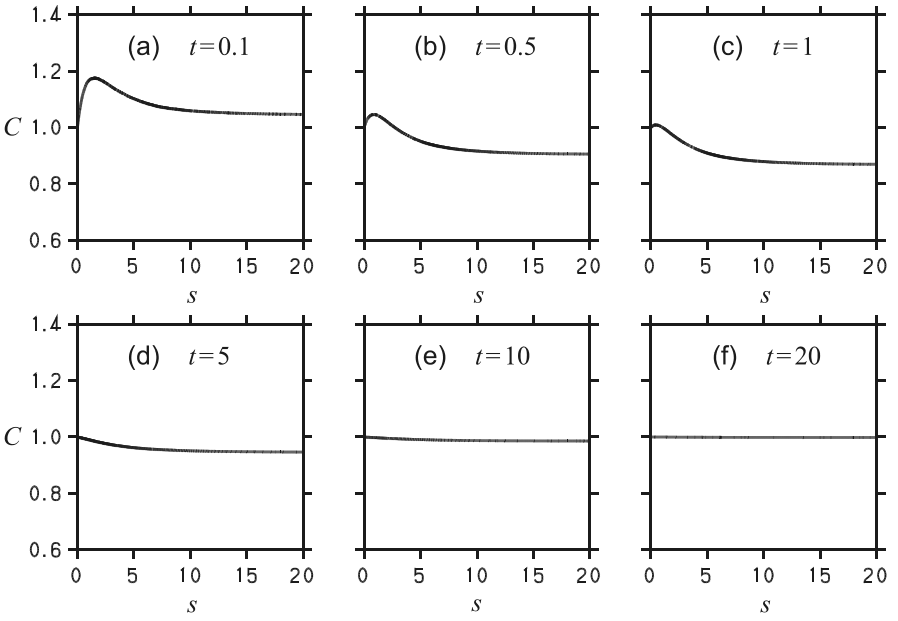



**Fig. 6.12.** The normalized correlation function (6.97) as a function of $s = m\,\Delta t$ in years for several fixed values $t = n\,\Delta t$ (in years).

Here, the value of $t$ is considered to be fixed, this means $C(t, s)$ is a function of $s$. Figure 6.12 shows $C(t, s)$ at the $t$ values considered in Fig. 6.10 and other $t$ values. The overall picture is that $C(t, s)$ becomes constant after a transitional stage. The reason for this is the lack of a decorrelation mechanism as given by the relaxation term $-v_{n-1} / \tau$ in Eq. (6.59b). The limit $C_\infty(t)$ of $C(t, s)$ for the case $s \to \infty$ is shown in dependence on $t$ in Fig. 6.13. It is interesting to see that $C_\infty(t) > 1$ for very small values of $t$. This behavior is caused by the fact that the fluctuations are initially equal to zero (see Fig. 6.11b). The generation of fluctuations then implies relatively large correlation values. In the next stage we see a decrease of the $C_\infty(t)$ values until the minimum at $t = 1.25$ is reached. This stage is characterized by a rapid distribution of the population PDF: see Fig. 6.10. This distribution of probability implies a correlation decrease (the same behavior is found for the evolution of a normal PDF with increasing variance). The reason that $C_\infty(t)$ does not continue to decay for increasing $t$ values is given by the equilibrium values $p = 0$ and $p = 1$ that confine the distribution of probability. Thus, the stage $t \geq 1.25$ is characterized by an increase of $C_\infty(t)$ until the value one is reached. This is the stage where the peaks close to the equilibrium values $p = 0$ and $p = 1$ develop, this means the correlation of population values has to increase again due to the trend of realizing the equilibrium values.

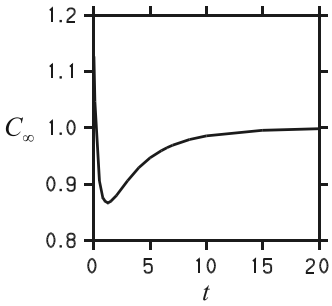



**Fig. 6.13.** The limit $C_\infty(t)$ of the standardized correlation function $C(t, s)$ for the case $s \to \infty$.

### *6.5.3 Model Application*

**Time to Extinction.** Let us calculate the time to extinction to illustrate the use of the stochastic population model. The time to extinction will be defined here to be the earliest time at which a realization is found in the interval $0 \le P_n \le L$, where $L = 10^{-8}$. The idea of using such a small $L$ value is to perform this calculation in a close agreement with the exact theoretical result. The influence of variations of $L$ on the time to extinction PDF is described in terms of exercise 6.5.5.

**Time to Extinction PDF.** The time to extinction is a random variable, which is different for every realization. To characterize this variable, its PDF $f(\tau)$ has been calculated. Here, $\tau$ refers to the sample space time to extinction in years. The PDF $f(\tau)$ was calculated at positions with a distance of 0.1 by using a filter size of 1, where $N = 10^6$ realizations were applied. The differences to the values used for the calculation of the population density PDF $f(p)$ arise from two facts. First, the range of data values is much larger for $f(\tau)$. Second, there are less sample values available for the calculation of $f(\tau)$ because only a part of all the realizations will approach the equilibrium value $p = 0$. The time to extinction PDF $f(\tau)$ is shown in Fig. 6.14. Here, the same model settings are used as for the population density PDF $f(p)$, except that variations of $\sigma$ are considered. The behavior of the time to extinction PDF can be explained by having again a look at Fig. 6.10. There are two stages. First, the number of population density values that are relatively close to $p = 0$ steadily increases in time because the population probability flows toward the equilibrium point $p = 0$. The increasing number of zero-population values in time is reflected by an increase of the time to extinction PDF with the characteristic arrival time $\tau$ at zero until the maximum is reached. The second stage shows a decrease of the time to extinction PDF with $\tau$ because of the lack of new population density values that approach $p = 0$. The development in the second stage takes place at a lower pace because it is not driven by a probability flow like the development in the first stage. Therefore, the time to extinction PDFs are skewed.

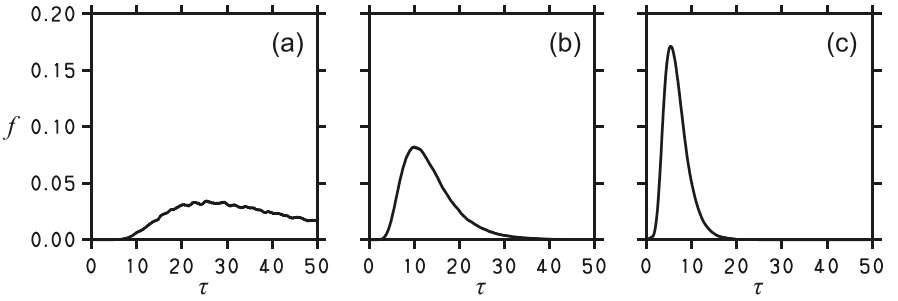



**Fig. 6.14.** The time to extinction PDF $f(\tau)$ implied by the population model (6.92). The values $\sigma = 1.5$, $\sigma = 2$, and $\sigma = 2.5$ are applied in (a), (b), and (c), respectively.

Figure 6.14 reveals that $\sigma$ has a significant influence on the time to extinction PDF: the higher the noise intensity $\sigma$, the smaller is the typical time to extinction. The value $\sigma = 2$ was chosen here as a reference value because the findings for this case are similar to the results of a population viability analysis for the endangered island fox (golden eagle predation is the primary cause of island fox mortality) (see Bakker at el. 2009). Why is the Fig. 6.14a PDF more noisy than the other curves? For this relatively low $\sigma$ value we have only 6.9% of all realizations that reach the equilibrium point $p = 0$. For $\sigma = 2$ and $\sigma = 2.5$, we have the case that 39.8% and 54.1% of the realizations reach the equilibrium point $p = 0$.

**Initial Population Effect.** It is interesting to consider the influence of the initial population density $P_0$ on the time to extinction PDF $f(\tau)$. This question is addressed in terms of Fig. 6.15 that shows the time to extinction PDF $f(\tau)$ for different initial population values $P_0$. It may be seen that the influence of $P_0$ on $f(\tau)$ is very limited. The PDF becomes more noisy with increasing $P_0$. The reason for this is given by the number of available realizations. For the cases considered, we have 39.8%, 18.5% and 5.9% of all realizations that are involved in the time to extinction PDF calculation shown in (a), (b), and (c), respectively.

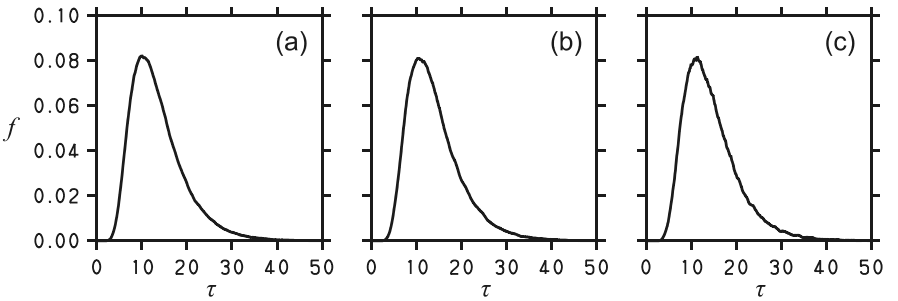



**Fig. 6.15.** The time to extinction PDF $f(\tau)$ is shown for $P_0 = 0.25$, $P_0 = 0.50$, and $P_0 = 0.75$ in (a), (b), and (c), respectively. All PDFs are obtained by using $\sigma = 2$.

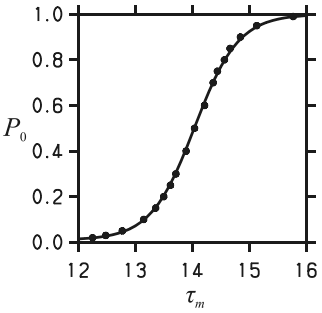



**Fig. 6.16.** The dependence of the mean time to extinction $\tau_m$ on the initial population $P_0$. The *dots* show the results obtained by using the population model (6.92). The *line* shows the approximation (6.98).

To see the effect of the initial population density $P_0$ more clearly, let us consider the effect of $P_0$ on the mean time to extinction $\tau_m$. Instead of considering $\tau_m$ as a function of $P_0$ we look here at the equivalent function $P_0(\tau_m)$ because $P_0(\tau_m)$ varies from zero to one. The variation $P_0(\tau_m)$ is presented in Fig. 6.16. The dependence $P_0(\tau_m)$ is plausible. The mean time to extinction $\tau_m$ increases with $P_0$ because it takes longer to realize the equilibrium value $p = 0$. For very small (very high) $P_0$, one has to expect that the $\tau_m$ values become very small (very high). The analytical approximation

$$P_0 = 0.01 + \frac{0.99}{1 + e^{-2.6(\tau_m - 14.04)}} \tag{6.98}$$

describes the variation of $P_0(\tau_m)$ very well.

## 6.6 Summary

The basic goal of this chapter was the explanation of how randomness should be involved in the deterministic difference equations considered in Chap. 5. For doing this we need knowledge about an appropriate structure of noise models and the influence of noise models on the statistical properties of a process considered. Finally, there is the question of how we can evaluate the suitability of a model for a stochastic process. Let us summarize the observations made in this chapter with regard to the latter questions.

**Noise Model.** The definition of an appropriate noise model requires two sorts of information: we have to define the PDF of noise values, and we have to define the scaling of the noise process with time. We did only consider here noise that is normally distributed. According to the discussions in Chap. 4, this is the most natural assumption for many problems, but this approach may turn out to be inappropriate for the modeling of variables that are non-negative or bounded by zero and one. Regarding the scaling with time we concluded that the consideration

of $\Delta W_n = (\Delta t)^{1/2}\ \varepsilon_n$ is the appropriate choice for scaled noise variables, where $\varepsilon_n$ is normally distributed. The reason for that can be seen by considering the diffusion model $x_n - x_{n-1} = D^{1/2}\ \Delta W_{n-1}$. This model provides a chance of $x_n$ that scales with $(\Delta t)^{1/2}$, which means that the variance of $x_n$ changes by a contribution that scales with $\Delta t$. Hence, after $n$ time steps we have a variance that is proportional to $n\ \Delta t$. This means that we calculate the variance correctly as a function of time $t = n\ \Delta t$. However, the use of $\Delta W_n = (\Delta t)^{1/2}\ \varepsilon_n$ for modeling the noise implies to deal with the problematic property of $\Delta W_n$ that

$$\left\langle \frac{\Delta W_n}{\Delta t}\frac{\Delta W_m}{\Delta t} \right\rangle = \begin{cases} \dfrac{1}{\Delta t} & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases}. \tag{6.99}$$

This relation means that the variance of $\Delta W_n / \Delta t$ goes to infinity as $\Delta t \to 0$.

**Effect of Noise on Model Statistics.** A general observation of the discussion of processes in this chapter is that the normality of noise does result in normally distributed process variables, provided that the equation considered is linear and the initial values are normally distributed. This conclusion does also apply to the linear equation system (6.101). In particular, we have to ask whether the use of $\Delta W_n$ with the property (6.99) means that we develop models with poor statistical properties. The suitability of using $\Delta W_n$ as scaled noise variables was justified by considering the diffusion equation

$$\frac{x_n - x_{n-1}}{\Delta t} = \sqrt{D}\,\frac{\Delta W_{n-1}}{\Delta t}. \tag{6.100}$$

It is correct that this model is a questionable model for the velocity (the left-hand side), but it was shown that the model represents a valid model for the particle position $x_n$. A better model is given by the Brownian motion model (6.59)

$$\frac{x_n - x_{n-1}}{\Delta t} = v_{n-1}, \tag{6.101a}$$

$$\frac{v_n - v_{n-1}}{\Delta t} = -\frac{1}{\tau}\left( v_{n-1} - \sqrt{D}\,\frac{\Delta W_{n-1}}{\Delta t} \right). \tag{6.101b}$$

This model represents a sound model for both the velocity $v_n$ and position $x_n$, but not for the particle acceleration (the left-hand side of Eq. (6.101b)). To have a good model for the particle position, velocity, and acceleration, one must consider a model where the derivative of the acceleration is driven by $\Delta W_{n-1} / \Delta t$ (see Heinz 2003). Corresponding conclusions apply to the population model (6.92),

$$\frac{P_n - P_{n-1}}{\Delta t} = P_{n-1}\left(1 - P_{n-1}\right)\left( \mu + \sigma\,\frac{\Delta W_{n-1}}{\Delta t} \right), \tag{6.102}$$

which represents an appropriate model for the variable $P_n$ considered.

**Evaluation of Stochastic Models.** The evaluation of the suitability of models for stochastic processes requires the evaluation of two sorts of model properties: the one-point statistics (PDF, mean, and variance), and correlations. The one-point statistics provide information about the typical value and range of variations of any variable considered. Correlations provide information about the characteristic lifetime of fluctuations, which is relevant to the modeling of processes. The evaluation of these properties is relatively simple as long as the difference equation considered is linear and the model parameters (the noise term and initial values) are normally distributed: we have then analytical expressions for the PDF and the correlations. The analysis of nonlinear equations is significantly more challenging. It is impossible to determine analytically the shape of the PDF, the equations for moments are unclosed due to the appearance of higher-order moments, and it is impossible to find an analytical expression for correlations. In the latter case, the statistical properties of processes considered have to be studied by simulations, as demonstrated with regard to population dynamics modeling in Sect. 6.5.

## 6.7 Exercises

**6.2.1** Consider the difference equation $y_n = a\,y_{n-1} + b + r\,\varepsilon_{n-1}$ discussed in Sect. 6.2.
a) Derive the difference equation for the mean $<y_n>$. Use this equation for the calculation of the equilibrium value of $<y_n>$.
b) Derive the difference equation for the variance $<\tilde{y}_n^{\,2}>$. Use this equation for the calculation of the equilibrium value of $<\tilde{y}_n^{\,2}>$.
c) Use the variance equation to explain why the equilibrium variance of $y_n$ is larger than the variance $r^2$ of the noise process $r\,\varepsilon_k$.

**6.2.2** Consider again the equation $y_n = a\,y_{n-1} + b + r\,\varepsilon_{n-1}$. The initial value is given by $y_0 = \varepsilon_0$. Assume that this equation is incorrectly solved by using for each $n = 1, 2, \ldots$ always the same set of random numbers $\varepsilon_n$.
a) Calculate analytically the mean value $<y_n>$ for this case.
b) Calculate analytically the standard deviation $<\tilde{y}_n^{\,2}>^{1/2}$ for this case.
c) Calculate analytically the correlation $<\tilde{y}_0\tilde{y}_n>$ for this case.
d) What is the difference between these results and the correct results for the case that $0 < a < 1$ and $n \to \infty$?

**6.2.3** Consider the equation $y_n = 0.2\,y_{n-1} + 0.1 + \varepsilon_{n-1}$, where $y_0$ is normally distributed with zero mean and standard deviation 0.1. The random variable $\varepsilon_{n-1}$ is normality distributed and characterized by $<\varepsilon_k> = 0$ and $<\varepsilon_k\,\varepsilon_m> = \delta_{km}$.

a) Graph three realizations of $y_n$.
b) Graph the numerically calculated mean $<y_n>$ for $N = 10^4$ realizations as a function of $n$.
c) Graph in the same figure the corresponding theoretical result.
d) Graph the numerically calculated standard deviation $<\tilde{y}_n^{\,2}>^{1/2}$ for $N = 10^4$ realizations as a function of $n$.
e) Graph in the same figure the corresponding theoretical result.

**6.2.4** Consider again the stochastic model considered in exercise 6.2.3.
a) Graph the numerically calculated filtered PDF at $n = 0$ for $N = 10^4$ realizations. Use a filter interval of 0.05.
b) Graph in the same figure the corresponding normal PDF at $n = 0$.
c) Graph the numerically calculated filtered PDF at $n = 1$ for $N = 10^4$ realizations. Use a filter interval of 0.05.
d) Graph in the same figure the corresponding normal PDF at $n = 1$.
e) Explain the reason for the difference between the PDF at $n = 0$ and the PDF at $n = 1$.

**6.2.5** Consider again the stochastic model considered in exercise 6.2.3.
a) Graph the numerically calculated normalized correlation function $C_0(n) = <\tilde{y}_0\tilde{y}_n> / <\tilde{y}_0^{\,2}>$ for $N = 10^4$ realizations.
b) Graph in the same figure the corresponding theoretical result.

**6.3.1** Apply Eq. (6.12) for $\tilde{y}_n$ to show that the correlations of the random walk model (6.28) are characterized by $<\tilde{y}_n(\tilde{y}_{n+m} - \tilde{y}_n)> = 0$.

**6.3.2** Consider the diffusion model $x_n - x_{n-1} = d\,\Delta t\ \varepsilon_{n-1}$. The random number $\varepsilon_{n-1}$ is normality distributed and characterized by $<\varepsilon_k> = 0$ and $<\varepsilon_k\ \varepsilon_m> = \delta_{km}$, and $d$ is a diffusion coefficient.
a) Calculate the variance of $x_n$.
b) Present the variance obtained as a function of time $t = n\,\Delta t$. Explain the problem related to the use of this variance.

**6.3.3** Show that the function

$$f(y) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{(y - y_0)^2}{2\sigma^2}\right\}.$$

approaches for $\sigma \to 0$ a delta function, $f(y) \to \delta(y - y_0)$.

**6.3.4** There is an instantaneous emission of a mass of 0.1 kg from a point source at a height of 10 m. The mean wind velocity in $x$ direction is 10 m / s. The diffusion coefficient is given by the mean wind velocity in $x$ direction multiplied with the height of 10 m.

a) Calculate the $x$ position at which the concentration has a maximum.
b) Calculate the maximum concentration at this position.

**6.3.5** Consider the diffusion model $y_n - y_{n-1} = D^{1/2}\,\Delta W_{n-1}$, which was discussed in Sect. 6.3. The old position is $y_{n-1} > 0$. The new position $y_n$ obtained by solving the diffusion equation is found to be beyond a totally reflecting boundary at $y = 0$, this means $y_n < 0$.
a) Calculate the mean velocity $v$ over the time step $\Delta t$.
b) Find the time $t_1$, which is the time required to reach the boundary. Find the time $t_2$, which is the remaining time over the time step $\Delta t$.
c) Calculate the new position $y_{n+1}$, which is obtained after moving with the velocity $v$ over the time $t_2$ in the positive $y$ direction.

**6.3.6** Consider the diffusion model $y_n - y_{n-1} = D^{1/2}\,\Delta W_{n-1}$ from Sect. 6.3. Assume that there is a totally absorbing boundary.
a) Show that the integral over the concentration (6.53) from zero to infinity is given by the following formula, which describes how much material is left in the accessible domain $0 \le y < \infty$,

$$\int_0^\infty C_n\,dy = \frac{M}{\sqrt{\pi}} \int_{-y_0/\sqrt{2\sigma_n^2}}^{y_0/\sqrt{2\sigma_n^2}} e^{-s^2}\,ds.$$

b) Consider the data of the diffusion problem described in exercise 6.3.4. Calculate the total amount of mass in the domain $0 \le y < \infty$ at $x = 10$ m. Hint: use the error function (see Sect. 4.3.2) to calculate the integral.

**6.3.7** Consider the diffusion model $y_n - y_{n-1} = D^{1/2}\,\Delta W_{n-1}$ from Sect. 6.3.
a) Derive an analytical formula for the concentration distribution that can be used for partially absorbing and partially reflecting boundaries. Show that this formula covers the cases of a totally reflecting boundary, of no boundary, and of a totally absorbing boundary.
b) Use this formula to find the position of the maximum and the maximum ground concentration for a boundary with 50% reflection.

**6.4.1** Prove the following identities:

a) $$(1-a^{n-1})^2 + (1-a^{n-2})^2 + \cdots + (1-a^2)^2 + (1-a)^2 = n - 2\frac{1-a^n}{1-a} + \frac{1-a^{2n}}{1-a^2}.$$

b) $$a^{n-1}(1-a^{n-1}) + a^{n-2}(1-a^{n-2}) + \cdots + a^2(1-a^2) + a(1-a) =$$
$$= \frac{1-a^n}{1-a} - \frac{1-a^{2n}}{1-a^2}.$$

**6.4.2** Prove Eqs. (6.72) for the statistics of the Brownian motion model:

a) $\left\langle \tilde{x}_n^{\,2} \right\rangle = \left\langle \tilde{x}_0^{\,2} \right\rangle + \tau^2 \left\langle \tilde{v}_0^{\,2} \right\rangle (1-a^n)^2 + D\left[ n\Delta t - 2\tau(1-a^n) + \tau \frac{1-a^{2n}}{1+a} \right].$

b) $\left\langle \tilde{x}_n \tilde{v}_n \right\rangle = \tau \left\langle \tilde{v}_0^{\,2} \right\rangle (1-a^n)\, a^n + D\left[ 1-a^n - \frac{1-a^{2n}}{1+a} \right].$

c) $\left\langle \tilde{v}_n^{\,2} \right\rangle = a^{2n} \left\langle \tilde{v}_0^{\,2} \right\rangle + \frac{D}{\tau} \frac{1-a^{2n}}{1+a}.$

Hint: use the identities given in exercise 6.4.1.

**6.4.3** Consider Eqs. (6.70) and (6.72) for means and variances of the Brownian motion model. For $n = 0$, these relations represent identities for the corresponding initial values.

a) Specify these relations for $n \geq 1$ for the case that the Brownian motion model is reduced to the diffusion model (6.63).

b) Calculate the initial mean and variance of the velocity by setting $n = 0$ in the relations for $n \geq 1$. This is a consistency condition if $\Delta t \to 0$.

c) Specify the relations from a) by using the initial values obtained in b).

**6.4.4** Consider the Brownian motion model (6.59).

a) Derive the discrete correlation function $<\tilde{v}_n \tilde{v}_{n+m}>$, where $n = 0, 1, \ldots$ and $m = 0, 1, \ldots$.

b) Set $t = n\,\Delta t$ and $s = m\,\Delta t$. Specify the correlation function obtained in a) to find the correlation function $<\tilde{v}(t)\tilde{v}(t+s)>$ in the limit $\Delta t \to 0$.

c) Integrate the normalized correlation function $<\tilde{v}(t)\tilde{v}(t+s)> / <\tilde{v}(t)^2>$ over s from zero to infinity.

d) What does this result mean regarding the relevance of $\tau$?

**6.4.5** Consider the Brownian motion model (6.59).

a) Prove the identity

$$(1-a^{n-1})(1-a^m a^{n-1}) + (1-a^{n-2})(1-a^m a^{n-2}) + \cdots +$$
$$+ (1-a^2)(1-a^m a^2) + (1-a)(1-a^m a) = n - (1+a^m)\frac{1-a^n}{1-a} + a^m \frac{1-a^{2n}}{1-a^2}.$$

b) Apply Eq. (6.71a) and the identity given in a) to derive the discrete correlation function $<\tilde{x}_n \tilde{x}_{n+m}>$, where $n = 0, 1, \ldots$ and $m = 0, 1, \ldots$.

c) Set $t = n\,\Delta t$ and $s = m\,\Delta t$. Specify the correlation function obtained in a) to find the correlation function $<\tilde{x}(t)\tilde{x}(t+s)>$ in the limit $\Delta t \to 0$.

d) Consider the correlation function $<\tilde{x}(t)\tilde{x}(t+s)>$ for $t \to \infty$.

e) Interpret the result obtained in d).

**6.5.1** Consider Eqs. (6.94) and (6.96) for the first-order and the second-order moments of the population density of the stochastic population model.
a) Specify these relations for the equilibrium case $n \to \infty$.
b) Specify the relations obtained in a) for the case that $\Delta t \to 0$.
c) Use the relations obtained in b) to write the standard deviation as a function of the mean.

**6.5.2** Consider the stochastic population model (6.92). Solve this equation numerically by following the settings described in Sect. 6.5.2.
a) Graph three realizations that apply the same $\sigma$.
b) Graph three realizations that apply $\sigma = 0.5$, $\sigma = 2$, and $\sigma = 7.5$. Use the same set of random numbers in all the three realizations.
c) Comment on the effect of $\sigma$.

**6.5.3** Continue with exercise 6.5.2.
a) Graph the mean and standard deviation that are shown in Fig. 6.11.
b) Calculate the so-called intensity of segregation, which is defined by $S_n = <\widetilde{P}_n^{\,2}> / [<P_n> (1 - <P_n>)]$.

**6.5.4** Consider Eqs. (6.94) and (6.96) for the first-order and the second-order moments of the population density of the stochastic population model.
a) Use the beta PDF to parametrize third- and fourth-order moments in Eqs. (6.94) and (6.96) in terms of first- and second-order moments. Hint: use the moment relation in exercise 4.3.10 for doing this.
b) Use these parametrizations in Eqs. (6.94) and (6.96) so that Eqs. (6.94) and (6.96) represent closed equations for first-order and second-order moments. Solve these equations by using a time step $\Delta t = 0.01$. Graph the mean population and standard deviation that follow from the solution of these equations as function of $n$.
c) Compare these curves with the exact mean and variance calculated in exercise 6.5.3. Discuss the suitability of using the beta PDF to close the moment equations.

**6.5.5** Consider the time to extinction PDF calculated from the stochastic population model in Sect. 6.5.3.
a) Graph the time to extinction PDF given in Fig. 6.15a by following the settings used in Sect. 6.5.3 (use $P_0 = 0.25$, $\mu = 1$, $\sigma = 2$, $\Delta t = 0.01$, and $L = 10^{-8}$).
b) Graph the corresponding time to extinction PDFs for the cases $L = 10^{-4}$ and $L = 10^{-2}$.
c) Explain why the influence of $L$ variations on the time to extinction PDF is reasonable.

# 7 Deterministic Evolution

Deterministic changes were discussed in Chap. 5 by analyzing several types of equations for changes (linear first-order equations, linear second-order equations, and nonlinear first-order equations). The goal of this discussion was to understand the nature of processes that are described by a given equation. In this chapter we ask how we can model any observed process in terms of equations. In particular, we are interested in an understanding of the laws that explain the development of any process. Is there a difference to the question considered in Chap. 5? A simple answer to this question is that an equation can describe several processes (see Fig. 5.4 that shows processes provided by a second-order equation), while the modeling of an observed process usually involves several equations (as given regarding the nonlinear and linear pendulum equations discussed in Chaps. 3 and 9). Hence, there is a difference to the approach used in Chap. 5. A more precise answer to this question is: The analysis of many processes enables the derivation of equations from observations. Such equations may be new equations. It may be also the case that such equations are so complicated that their use requires systematic simplifications. Thus, the analysis of observed processes does often end up with new equations and a hierarchy of simpler equations that can be applied under specific conditions – which is different to the approach of Chap. 5 to work with one given equation. The approach applied to address the modeling of processes is the following: First, we will focus on basic processes (those that have an increasing intensity, processes that level off, oscillating processes). Second, we only consider single variable problems here (the modeling of processes involving several variables will be addressed in Chap. 9). Third, we consider only two sorts of processes: mechanical processes (for which we have well established laws) and population ecology processes (for which we have an ongoing debate about the existence of general laws).

S. Heinz, *Mathematical Modeling*, DOI 10.1007/978-3-642-20311-4_7,

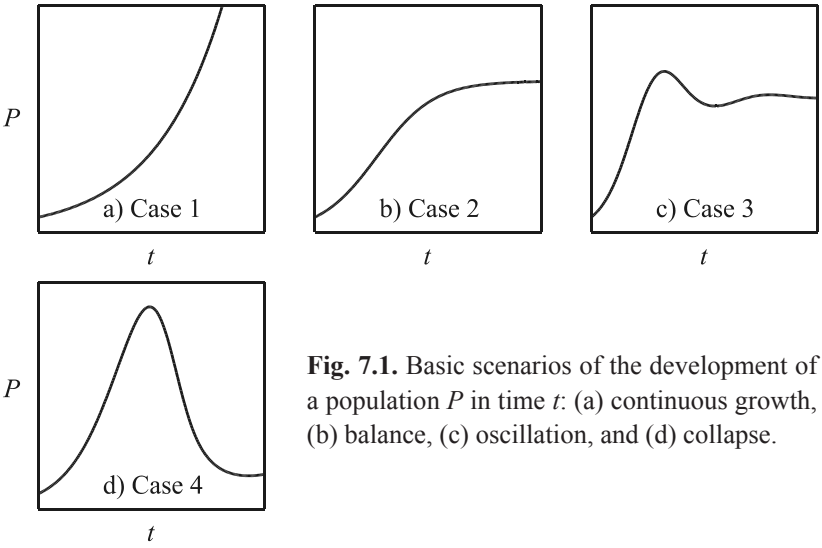



**Fig. 7.1.** Basic scenarios of the development of a population $P$ in time $t$: (a) continuous growth, (b) balance, (c) oscillation, and (d) collapse.

This chapter is organized in the following way. Section 7.1 will describe basic scenarios of the development of processes and characteristic questions regarding the modeling of such processes. The modeling of mechanical processes with and without oscillations are addressed in Sects. 7.2, and 7.3, respectively. The modeling of processes in population ecology with and without oscillations is the concern of Sects. 7.4 and 7.5, respectively. Section 7.6 will summarize the features of laws in mechanics and population ecology considered here.

## 7.1 Motivation

**Population Dynamics.** Processes that we have to consider usually reveal a variety of scenarios. An example for the wide range of process variations is given in Fig. 7.1 that shows possible developments of a population density (see the more detailed discussion in Sect. 7.5.1). Case 1 illustrates an unbounded population growth under optimal living conditions. Case 2 illustrates the development of a final equilibrium that is given by the maximum population that can be supported by the environment. Case 3 illustrates damped oscillations about an equilibrium state, which can be related to a delayed response of the supporting environment to the growing population (see the explanations in Sect. 7.5.1). Case 4 illustrates a breakdown of the environmental system that causes a collapse of the population (collapse means a dramatic reduction of past population values). The difference between these four cases is given by the interaction between the population and

the supporting system ranging from a zero interaction (case 1) to a strong coupling (case 4). Such a variety of interactions makes it rather unlikely that all these cases can be modeled on the basis of only one model: at least the cases that do and do not involve oscillations will require different models (as will be shown below). On the other hand, a helpful observation is that similar basic scenarios can be found for other processes. With regard to mechanical processes, for example, we often find processes that are characterized by the cases 2 and 3. Examples for such processes are given by the balance of temperature differences and motions of a spring-mass system, respectively.

**Questions Considered.** The development of mathematical models for the cases illustrated in Fig. 7.1 (and for many other cases) leads to questions like:

- How can we derive model equations?
- Which cases can we cover on the basis of one equation?
- Can we use the same equations for similar processes in different applications?

From a more general point of view there are questions like:

- What are the characteristic properties of a law?
- Are there laws of mechanics?
- Are there laws of population ecology?

These and other questions will be addressed in the following based on the consideration of basic mechanical processes and population ecology processes

## 7.2 Heat and Mass Transfer: Balance

Let us discuss first a relatively simple but very important case: the modeling of the distribution of heat or mass. The basic problem is usually the modeling of the transition from any initial state (the emission of heat or mass from any source) to an equilibrium state (at which heat or mass distributions do not change anymore). Such processes may be well described by first-order linear differential equations.

### *7.2.1 Heat and Mass Transfer Models*

**Mass Transfer Model.** How can we model the transfer of mass? Let us say $y$ is the mass of a substance in a tank. We do not consider any spatial distribution of mass, but the mass is characterized by one global value. Then, the change of $y$ in time $t$ can be characterized by the equation

$$\frac{dy}{dt} = \left(\frac{dy}{dt}\right)_{in} - \left(\frac{dy}{dt}\right)_{out}. \tag{7.1}$$

We consider a differential equation because the process considered is a continuous function of $t$. According to this relation, $y$ changes due to two contributions: the inflow rate $(dy/dt)_{in}$, which increases $y$, and the outflow rate $(dy/dt)_{out}$, which decreases $y$. To find an equation for $y$ we have to relate the inflow and outflow rates to $y$. Regarding the outflow rate, it is a reasonable assumption to assume that $(dy/dt)_{out}$ is proportional to $y$, which means that

$$\left(\frac{dy}{dt}\right)_{out} = \frac{y}{T}. \tag{7.2}$$

$T$ is a time scale that characterizes the outflow. The need to involve a time scale in this relation arises from the requirement to design a dimensionally correct equation. It is also reasonable to consider a similar structure for $(dy/dt)_{in}$,

$$\left(\frac{dy}{dt}\right)_{in} = \frac{y_e}{T}. \tag{7.3}$$

Here, $y_e$ is a variable that has the same dimension as $y$. The combination of Eqs. (7.1), (7.2), and (7.3) then leads to the equation

$$\frac{dy}{dt} = -\frac{y - y_e}{T}. \tag{7.4}$$

This equation explains the physical relevance of $y_e$: $y_e$ is an equilibrium solution because there is no change of $y$ anymore if $y = y_e$. The model parameters $y_e$ and $T$ have to be provided to have a model for $y$. Both $y_e$ and $T$ are assumed to be independent of $y$, but these parameters can be functions of $t$.

**Heat Transfer Mechanisms.** Let us consider heat transfer to understand which sort of equation is needed to describe this process. It is important to know that there is not just one heat transfer mechanism, but there are several processes that may take place. A first mechanism is convection that takes place through diffusion (the random Brownian motion of individual particles in a fluid) and advection (the transport of heat by the larger-scale motions in the fluid). This means, convection is the transport of heat by the actual movement of warmed matter (for example, the movement of heat that leaves a hot cup of coffee as the currents of steam and air rise). A second mechanism is conduction, which is the transfer and distribution of heat energy from atom to atom within a substance. For example, a spoon in a cup of hot soup becomes warmer because the heat from the soup is conducted along the spoon. A third mechanism is thermal radiation, which is the transfer of heat energy through empty space due to electromagnetic waves. An example is given by sunlight that is radiated through space to the Earth. In the following we will only consider the mathematical modeling of convection. For doing this we consider the heat transfer in still air where the temperature difference is relatively small (about 20°K – 30°K).

**Heat Transfer Model.** To explain the assumptions that underlie Newton's Law of Cooling let us consider the following equation for the transfer of temperature $y$,

$$\frac{\partial y}{\partial t} + U\frac{\partial y}{\partial x} = 0. \tag{7.5}$$

This equation represents a partial differential equation (readers who are unfamiliar with such equations can skip the following explanation of how Eq. (7.7) can be derived). Equation (7.5) is the result of simplifying a more general temperature equation that is derived in Sect. 10.5.2. The underlying idea of Eq. (7.5) is the assumption that a change of $y$ in time is possible through convective transport along an axis $x$ with a characteristic velocity $U$. Next, we assume that the spatial gradient $\partial y / \partial x$ can be approximated by $(y - y_e) / \Delta x$. Here, $y_e$ is a characteristic external (ambient) temperature, and $\Delta x$ characterizes the spatial scale of the heat difference. With this assumption, Eq. (7.5) reads

$$\frac{dy}{dt} = -U\frac{y - y_e}{\Delta x}. \tag{7.6}$$

In this equation, the partial derivative $\partial y / \partial t$ was replaced by the total derivative $dy / dt$ because the derivative by $t$ is the only derivative that is left here. By introducing now the characteristic time scale $T = \Delta x / U$ we find

$$\frac{dy}{dt} = -\frac{y - y_e}{T}. \tag{7.7}$$

This equation represents Newton's Law of Cooling. It assumes that the change of $y$ in time is controlled by the difference between $y$ and the external temperature $y_e$. The validity of the negative sign can be explained in the following way: For the case $y > y_e$, we find that $y$ decreases until $y$ becomes equal to $y_e$. Vice versa, for the case $y < y_e$, $y$ will increase until $y$ becomes equal to $y_e$. Equation (7.7) has the same structure as Eq. (7.4) for the conservation of mass. We assume again that $y_e$ and $T$ are independent of $y$, but they can be functions of $t$.

### *7.2.2 First-Order Linear Differential Equations*

Next, let us consider the solution of Eq. (7.7). This question will be only considered for the relevant case that the model parameters $y_e$ and $T$ are constant (techniques that allow the solution of Eq. (7.7) for the case that $y_e$ is a function of $t$ are described, for example, by Boyce & DiPrima 2009). The analytical solution of Eq. (7.7) was already obtained in Sect. 5.5.2 based on the corresponding difference equation. The solution of Eq. (7.7) as a differential equation will be shown in the following to illustrate the approach for solving differential equations.

**Solution.** Equation (7.7) can be solved by means of the method of separation of variables. For applying this method we write Eq. (7.7) such that the variables $y$ and $t$ are separated,

$$\frac{dy}{y - y_e} = -\frac{dt}{T}. \tag{7.8}$$

We can integrate this relation formally,

$$\int \frac{dy}{y - y_e} = -\int \frac{dt}{T} + C, \tag{7.9}$$

where $C$ is any unknown constant. Next, we perform the integration on both sides,

$$\ln | y - y_e | = -\frac{t}{T} + C. \tag{7.10}$$

By taking both sides as exponents of an exponential function we obtain

$$| y - y_e | = e^C e^{-t/T}. \tag{7.11}$$

This relation can be also written

$$y - y_e = \pm e^C e^{-t/T} = c\, e^{-t/T}, \tag{7.12}$$

where the constant $\pm e^C$ was replaced by another unknown constant $c$ to simplify the notation. The constant $c$ be determined by specifying Eq. (7.12) for $t = 0$,

$$y_0 - y_e = c. \tag{7.13}$$

Here, $y_0$ refers to the initial value of $y$ at $t = 0$. By using this expression for $c$ in Eq. (7.12) we find

$$y = y_e + e^{-t/T}(y_0 - y_e). \tag{7.14}$$

This expression recovers the result (5.120), which was obtained from the corresponding difference equation.

**Solution Check.** It is always helpful to check the validity of solutions obtained for differential equations. This requires, first, evidence that Eq. (7.14) provides the correct initial value, and, second, evidence that Eq. (7.14) indeed satisfies the differential equation (7.7). The correctness of the initial value provided by Eq. (7.14) can be seen by setting $t = 0$ in Eq. (7.14),

$$y(t = 0) = y_0. \tag{7.15}$$

To see that Eq. (7.14) indeed satisfies Eq. (7.7) we calculate the derivative of $y(t)$,

$$\frac{dy}{dt} = -\frac{1}{T} e^{-t/T}(y_0 - y_e) = -\frac{y - y_e}{T}. \tag{7.16}$$

The last expression is obtained by replacing the exponential function according to Eq. (7.14). Hence, Eq. (7.14) does indeed solve the differential equation (7.7).

### *7.2.3 Time of Death*

**Problem.** Let us consider an example to illustrate the use of Eq. (7.14) for the calculation of heat transfer. A police report provides the following facts: Police arrived at the scene of a murder at 8 a.m. They immediately took and recorded the temperature of the corpse, which was 33ºC, and thoroughly inspected the area. By the time they finished the inspection, it was 10 a.m. They again took the temperature of the corpse, which had dropped to 29ºC, and had the corpse sent to the morgue. The temperature at the crime scene had remained steady at 23ºC. We suppose that the corpse temperature obeys Newton's Law of Cooling. What is the time of death by assuming a normal body temperature of 37ºC at this time?

**Solution.** To address this problem, we consider Eq. (7.14) such that we account for an initial time $t_0$, which we try to calculate,

$$y = y_e + e^{-(t-t_0)/T}(y_0 - y_e). \tag{7.17}$$

The setting $t = t_0$ shows that $y(t_0) = y_0$. By taking the derivative of $y$ we can prove that Eq. (7.17) is also a solution of the differential equation (7.7),

$$\frac{dy}{dt} = -\frac{1}{T}e^{-(t-t_0)/T}(y_0 - y_e) = -\frac{y - y_e}{T}, \tag{7.18}$$

where Eq. (7.17) was used to obtain the last expression. The initial temperature $y_0 = 37$ºC, and the external temperature is $y_e = 23$ºC. Hence, Eq. (7.17) reads

$$y = 23 + 14e^{-(t-t_0)/T}. \tag{7.19}$$

We know that the temperature was 33ºC at 8 a.m., and 29ºC at 10 a.m., this means we have the relations

$$33 = 23 + 14e^{-(8-t_0)/T}, \qquad 29 = 23 + 14e^{-(10-t_0)/T}. \tag{7.20}$$

The latter two conditions can be also written

$$-(8 - t_0) = T\ln\left(\frac{33-23}{14}\right), \qquad -(10 - t_0) = T\ln\left(\frac{29-23}{14}\right). \tag{7.21}$$

By equating both implied expressions for $t_0$ we obtain

$$8 + T\ln(5/7) = 10 + T\ln(3/7). \tag{7.22}$$

Hence, $T$ is given by

$$T = \frac{2}{\ln(5/7) - \ln(3/7)} = \frac{2}{\ln(5/3)} = 3.92. \tag{7.23}$$

We can use this value in Eq. (7.21) to find

$$t_0 = 8 + 3.92 \ln(5/7) = 6.68. \tag{7.24}$$

Hence, the time of death was 6:41 a.m. The use of the values obtained for $T$ and $t_0$ in Eq. (7.19) provides the temperature function

$$y = 23 + 14 e^{-(t-6.68)/3.92}. \tag{7.25}$$

This formula is correct because it provides the correct temperature values 37ºC at $t_0 = 6.68$, 33ºC at 8 a.m., and 29ºC at 10 a.m.

### *7.2.4 Contamination of Lakes*

**Problem.** Let us consider a second example to illustrate the application of Eq. (7.14) to the calculation of mass transfer (Rainey 1967, Boyce & DiPrima 2009). A lake has a constant volume $V$ (measured in km$^3$) containing at time $t$ a mass $Q(t)$ (in kg) of pollutant. The initial mass at $t = 0$ is denoted by $Q_0$. The pollutant is evenly distributed throughout the lake with a concentration $c(t) = Q(t) / V$. We assume that water containing a constant concentration $c_{in}$ of pollutant enters the lake at a rate $r$ (in km$^3$/year), and that water leaves the lake at the same rate. Let us consider the following questions:

(1) What is the mass $Q(t)$ at any time $t$?

(2) If the addition of pollutants to the lake is terminated ($c_{in} = 0$), which time interval $\tau$ must elapse such that the mass $Q$ of pollutants is reduced to $Q_0 / \alpha$?

(3) Regarding the data given in Table 7.1, what is the time $\tau$ necessary to reduce the contamination of each of these lakes to 10% of the original value?

**Solution.** To find the answers to the latter questions we have to derive first the differential equation for $Q(t)$. The dimensions involved imply that the change of $Q(t)$ in time $t$ is determined by the flow rate $r$ multiplied by a concentration $c$. In particular, we have a balance between the inflow and outflow,

$$\frac{dQ}{dt} = (rc)_{\text{inflow}} - (rc)_{\text{outflow}}. \tag{7.26}$$

The rate $r$ and inflow concentration $c_{in}$ are constants. The outflow concentration is given by the actual pollutant concentration $c(t) = Q(t) / V$. Hence, Eq. (7.26) reads

$$\frac{dQ}{dt} = r c_{in} - r\frac{Q}{V} = -\frac{r}{V}(Q - V c_{in}). \tag{7.27}$$

**Table 7.1** Volume and flow data for the Great Lakes (Rainey 1967).

| Lake | $V$ (km$^3$) | $r$ (km$^3$/year) | $\tau$ (years) |
|---|---|---|---|
| Superior | 12,200 | 65.2 | 430.9 |
| Michigan | 4,900 | 158.0 | 71.4 |
| Erie | 460 | 175.0 | 6.1 |
| Ontario | 1,600 | 209.0 | 17.6 |

This equation corresponds to Eq. (7.7): the variables $Q$, $V\,c_{in}$, and $V/r$ used here correspond to the variables $y$, $y_e$, and $T$ used in Eq. (7.7), respectively. According to Eq. (7.14) the solution of Eq. (7.27) is then given by

$$Q = V\,c_{in} + e^{-rt/V}\left(Q_0 - V\,c_{in}\right), \tag{7.28}$$

which provides the answer to question (1). To find the answer to question (2) we apply $c_{in} = 0$ and use the condition

$$\frac{Q_0}{\alpha} = Q_0\,e^{-r\tau/V}. \tag{7.29}$$

The latter condition implies

$$\ln(1/\alpha) = -\frac{r\tau}{V}. \tag{7.30}$$

Hence, the answer to question (2) is given by

$$\tau = \frac{V}{r}\ln\alpha. \tag{7.31}$$

The answer to question (3) is given in Table 7.1, where $\alpha = 10$ is used.

## 7.3 Newton's Laws of Motion: Oscillations

Next, let us consider how another basic process can be modeled: damped or undamped oscillations between two equilibrium states. The basis for the modeling of this process is given by Newton's Laws of Motion. Although the focus in this chapter is on the discussion of modeling principles for single variable processes, Newton's Laws of Motion will be presented here for several variables in order to prepare the application of Newton's Laws to several variable cases in Chap. 9. Due to the use of vector notation, this approach does not add technical difficulties. The resulting equations will be used in this chapter only for single-variable cases.

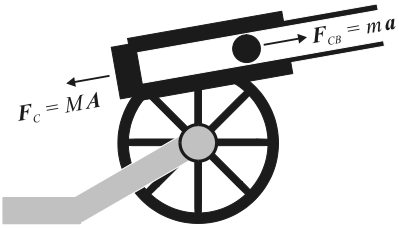



**Fig. 7.2.** An illustration of Newton's Third Law. The cannon has a mass $M$ and acceleration $\boldsymbol{A}$, and the cannon ball has a mass $m$ and acceleration $\boldsymbol{a}$. The forces $\boldsymbol{F}_{CB}$ and $\boldsymbol{F}_C$ of the cannon ball and cannon have the same magnitudes but opposite directions.

### *7.3.1 Newton's Laws of Motion*

**Newton's Laws of Motion.** Sir Isaac Newton developed in his famous book "Philosophiæ Naturalis Principia Mathematica" (1687), Latin for "Mathematical Principles of Natural Philosophy", a theory for the forces acting on a macroscopic body and the motion of that body. Newton's theory explains the motion of many physical objects and systems. For example, Newton demonstrated that these Laws of Motion, combined with his Law of Universal Gravitation, explain Kepler's Laws of Planetary Motion. Newton's Laws were verified by experiment and observation, and they are excellent approximations at the scales and speeds of everyday life. Newton's Laws of Motion, together with his Law of Universal Gravitation and the related developments of calculus, provided for the first time a unified quantitative explanation for a wide range of physical phenomena. Thus, Newton is seen by many scholars to be one of the most influential people in human history. Newton's Laws of Motion can be summarized by the following three laws:

- First Law (the law of inertia): An object will remain at rest or in uniform motion in a straight line if it is not affected by an external force.
- Second Law (the "force equals mass times acceleration" law): A body that is affected by a force $\boldsymbol{F}$ experiences an acceleration $\boldsymbol{a}$ that is related to the force by $\boldsymbol{F} = m\,\boldsymbol{a}$, where $m$ is the mass of the body.
- Third Law (the action-reaction law)**:** For every action, there is an equal and opposite reaction.

The bold symbols $\boldsymbol{F}$ and $\boldsymbol{a}$ refer to three-dimensional vectors, i.e., $\boldsymbol{F} = (F_1, F_2, F_3)$ and $\boldsymbol{a} = (a_1, a_2, a_3)$.

**Newton's Third Law.** Newton's Third Law means that all forces occur in equal but oppositely directed pairs: whenever a first body exerts a force $\boldsymbol{F}$ on a second body, the second body exerts a force $-\boldsymbol{F}$ on the first body. $\boldsymbol{F}$ and $-\boldsymbol{F}$ are equal in magnitude and opposite in direction. Newton used the Third Law to derive the Law of Conservation of Momentum (actually, conservation of momentum is the more fundamental idea). Newton's Third Law can be illustrated by considering a cannon that fires a cannon ball (see Fig. 7.2). The force $\boldsymbol{F}_{CB}$ that acts on the cannon ball is

given by $\boldsymbol{F}_{CB} = m\,\boldsymbol{a}$, where $\boldsymbol{a}$ is the cannon ball acceleration. The force $\boldsymbol{F}_C$ that acts on the cannon is given by $\boldsymbol{F}_C = M\,\boldsymbol{A}$, where $M$ is the mass of the cannon and $\boldsymbol{A}$ is the cannon acceleration. The two forces have the same magnitudes but opposite directions. Why does the cannon ball shoot out so far, whereas the cannon itself is only kicked back a little bit? Because $M >> m$, which implies that $|\boldsymbol{a}| >> |\boldsymbol{A}|$ due to $m\,|\boldsymbol{a}| = M\,|\boldsymbol{A}|$. Another example for Newton's Third Law is given by the flight of a balloon: if air is rushing out of a balloon, the reaction is that the balloon is forced away. Newton's Third Law is also the reason why aircraft can fly (see Chap. 3). The air is deflected downward by the action of the airfoil, and in reaction the wing is pushed upward by the lift force.

**Newton's Second Law.** Newton's Second Law provides an explanation for the relation between a force acting on a body and the motion of that body. In terms of the definition $\boldsymbol{a} = d\boldsymbol{v}/dt$ of acceleration, Newton's Second Law can also be written

$$\frac{d\boldsymbol{v}}{dt} = \frac{\boldsymbol{F}}{m}, \tag{7.32}$$

or, by using the definition $\boldsymbol{v} = d\boldsymbol{x}/dt$ of velocity, Newton's Second Law reads

$$\frac{d^2\boldsymbol{x}}{dt^2} = \frac{\boldsymbol{F}}{m}. \tag{7.33}$$

Thus, for a given force $\boldsymbol{F} = \boldsymbol{F}(d\boldsymbol{x}/dt, \boldsymbol{x}, t)$, Newton's Second Law provides a differential equation of second order for the calculation of the position $\boldsymbol{x}(t)$ of a body. Obviously, the velocity $\boldsymbol{v} = d\boldsymbol{x}/dt$ can be calculated if $\boldsymbol{x}(t)$ is known. Equation (7.33) provides the basis for solving many everyday life problems. However, this equation cannot be applied if the mass $m$ is changing, or if the object considered is traveling with a velocity close to the speed of light. Equation (7.33) does also not apply on the very small scale of atoms where quantum mechanics must be used.

**Newton's First Law.** Newton's First Law is a consequence of Newton's Second Law for the case that $\boldsymbol{F} = 0$. Equation (7.32) implies for this case

$$\frac{d\boldsymbol{v}}{dt} = 0, \tag{7.34}$$

which means that the velocity vector has to be constant,

$$\boldsymbol{v} = \text{constant}. \tag{7.35}$$

This writing means that both the magnitude and direction of the velocity vector are constant. Therefore, an object will remain at rest or in uniform motion in a straight line.

**Spring-Mass System.** Vibrations represent a basic feature of many processes. There are mechanical vibrations (pendulum), electrical vibrations (the electrical current in a simple series circuit), electro-mechanical vibrations (microphone), optical vibrations (laser), biological vibrations (Lotka–Volterra equation), climatic

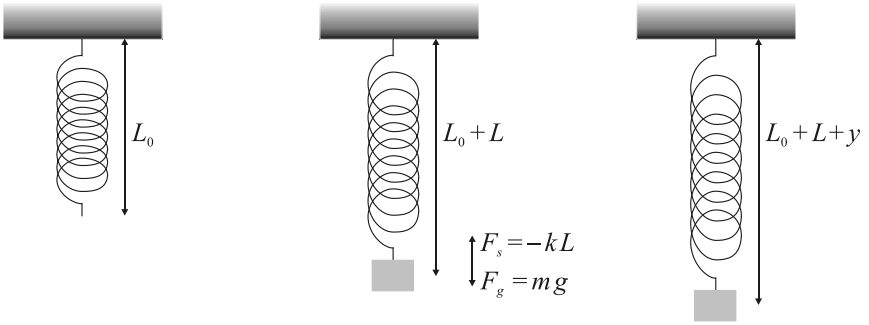



**Fig. 7.3.** An illustration of a spring-mass system.

oscillations (El Niño-Southern Oscillation), and chemical oscillations (Belousov–Zhabotinsky reaction). Let us consider a spring-mass system as a simple example for such oscillations and the use of Newton's Second Law for the calculation of the motion of bodies. The vertical spring-mass system is illustrated in Fig. 7.3. The original length of the spring without mass is $L_0$. The addition of a mass $m$ causes an elongation $L$ of the spring in the downward direction, which is considered to be the positive direction. Two forces act at the point where the mass is attached to the spring: the gravity force $F_g = m\,g$, $g$ refers to the gravity acceleration, and a spring force $F_s$ acting to restore the spring to its natural position. The spring force can be described by Hooke's Law if the elongation $L$ is relatively small. For this case, the spring force is proportional to the elongation, $F_s = -\,k\,L$. Here, $k$ is a positive constant, which is called the spring constant. A negative sign appears in $F_s = -\,k\,L$ because the spring force acts in the negative upward direction. The mass is in an equilibrium, which means that the mass does not move. The condition for this equilibrium is that the two forces balance each other, which means that they add to zero,

$$m\,g - k\,L = 0. \tag{7.36}$$

To study the motion of the spring system we use Newton's Second Law (7.33),

$$\frac{d^2 y}{dt^2} = \frac{F}{m}. \tag{7.37}$$

Here, $y(t)$ is the displacement of the mass from its equilibrium position at time $t$. According to our assumption above on $L_0$ and $L$, $y$ is positive in the downward direction. $F(t)$ represents the sum of all forces that act on the mass. In particular, $F(t)$ is assumed to be given by

$$F = F_g + F_s + F_d. \tag{7.38}$$

The three forces are related to $dy/dt$ and $y$ by the expressions

$$F_g = m\,g, \tag{7.39a}$$

$$F_s = -k\,(L + y), \tag{7.39b}$$

$$F_d = -\gamma \frac{dy}{dt}. \tag{7.39c}$$

The gravity force $F_g$ is unaffected by the displacement $y$ of the mass. The expression for the spring force $F_s$ is extended here by accounting for the varying distance $y$ in addition to the equilibrium elongation $L$ considered above, i.e., the spring force is assumed to be proportional to the total elongation $L + y$. As as-sumed above, we consider the elongation $L + y$ to be sufficiently small such that Hooke's Law can be applied. In addition to $F_g$ and $F_s$ we do also consider here a damping force $F_d$. This force acts opposite to the direction of motion of the mass: this force reduces the velocity $dy/dt$. Here, $\gamma$ is a positive constant that is called the damping constant. The structure of the damping force corresponds to Stokes' Law (see Sect. 3.3.3). The damping force may arise from the resistance from the air in which the mass moves. Interestingly, the assumption of a damping force that is linear in the velocity $dy/dt$ is only one reasonable assumption among several possible choices. In general, there can be also other forces that contribute to the total force $F(t)$, for example, any external force that implies forced vibrations. The combination of Newton's Second Law (7.33) with these assumptions about the forces involved then results in the following equation for the displacement $y(t)$,

$$\frac{d^2y}{dt^2} = \frac{1}{m}\left\{mg - k(L + y) - \gamma \frac{dy}{dt}\right\}. \tag{7.40}$$

We multiply this equation with $m$ and use the equilibrium condition $m\,g = k\,L$,

$$m\frac{d^2y}{dt^2} + \gamma \frac{dy}{dt} + k\,y = 0. \tag{7.41}$$

This equation represents a linear homogeneous differential equation of second order for the calculation of the displacement $y(t)$. From a more general point of view, Eq. (7.41) represents the equation for a damped harmonic oscillator. Such oscillator equations occur in a diverse range of disciplines, for example, in control engineering, mechanical engineering and electrical engineering.

**Undamped Harmonic Oscillator Solution.** How can we solve the damped harmonic oscillator equation (7.41)? It turns out that there is no simple answer to this question. To see the structure of solutions, let us consider a relatively simple case given by the undamped harmonic oscillator equation

$$\frac{d^2y}{dt^2} + \frac{k}{m}y = 0. \tag{7.42}$$

Due to dimensional reasons, the constant $m/k$ represents a squared time. Thus, we introduce the characteristic time scale $T = (m/k)^{1/2}$, such that (7.42) can be written

$$\frac{d^2y}{dt^2} + \frac{y}{T^2} = 0. \tag{7.43}$$

A solution to this equation is given by any function $y(t)$ having a second-order derivative $y''(t)$ that is proportional to $y(t)$. Exponential functions are characterized by the property $y''(t) \sim y(t)$. A closer look at Eq. (7.43) reveals that one possible solution is given by

$$y_1(t) = e^{it/T}. \tag{7.44a}$$

Here, $i$ is the imaginary unit defined by $i^2 = -1$. Another possible solution is

$$y_2(t) = e^{-it/T}. \tag{7.44b}$$

The validity of these two solutions can be proven by calculating the second-order derivatives of $y_1$ and $y_2$. A further consideration of this question reveals that any linear combination

$$y_3(t) = c_1 e^{it/T} + c_2 e^{-it/T} \tag{7.44c}$$

is also a solution to Eq. (7.43), where $c_1$ and $c_2$ are any constants (the rewriting of the solution (7.44c) in terms of real-valued functions will be shown below). The observations made here lead to two relevant conclusions: First, exponential functions may provide solutions for a linear, homogeneous, second-order differential equation with constant coefficients. Second, the general solution can be a superposition of two exponential solutions. Consequently, the calculation of solutions for the damped harmonic oscillator equation (7.41) is not very simple: it requires a detailed analysis that will be presented in Sect. 7.3.2.

### *7.3.2 Second-Order Linear Differential Equations*

**Differential Equation.** To find the solution of the damped harmonic oscillator equation (7.41) let us consider a linear, homogeneous, second-order differential equation with constant coefficients $a$, $b$, and $c$,

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + c\,y = 0. \tag{7.45}$$

This equation corresponds to the oscillator equation (7.41): the difference is that the positive constants $m$, $\gamma$, and $k$ in Eq. (7.41) are replaced here by any (positive or negative) constants $a$, $b$, and $c$, respectively. The solution of Eq. (7.45) depends on two constants that are produced by the two integrations required to solve this

equation. To determine these two constants, we consider conditions for the initial value of $y$ and the initial value of $dy/dt$,

$$y(0) = y_0, \tag{7.46a}$$

$$y'(0) = y'_0. \tag{7.46b}$$

Here, $y_0$ and $y'_0$ are considered to be known parameters.

**General Solution.** To find conditions for the existence of exponential solutions we assume a solution of the form

$$y(t) = e^{rt}, \tag{7.47}$$

where $r$ is any unknown constant. By using this assumption in Eq. (7.45) we find

$$(ar^2 + br + c)e^{rt} = 0. \tag{7.48}$$

This condition is satisfied for all $t$ if the parameter $r$ satisfies the condition

$$ar^2 + br + c = 0. \tag{7.49}$$

The latter equation is called the characteristic equation of the differential equation (7.45). The two roots $r_1$ and $r_2$ of the characteristic equation, which are called the eigenvalues, are given by

$$r_1 = \frac{1}{2a}\left[-b + \sqrt{b^2 - 4ac}\right], \qquad r_2 = \frac{1}{2a}\left[-b - \sqrt{b^2 - 4ac}\right]. \tag{7.50}$$

Another appropriate way to write $r_1$ and $r_2$, which will be used for the presentation of solutions below, is given by

$$r_1 = r_S + r_D, \qquad r_2 = r_S - r_D, \tag{7.51}$$

where $r_S$ and $r_D$ are given by

$$r_S = \frac{r_1 + r_2}{2} = -\frac{b}{2a}, \qquad r_D = \frac{r_1 - r_2}{2} = \frac{1}{2a}\sqrt{b^2 - 4ac}. \tag{7.52}$$

The general solution of Eq. (7.45) represents a linear superposition of the two possible solutions $\exp(r_1 t)$ and $\exp(r_2 t)$,

$$y(t) = c_1 e^{r_1 t} + c_2 e^{r_2 t}. \tag{7.53}$$

The two unknown constants $c_1$ and $c_2$ can be calculated by the constraint that $y(t)$ has to satisfy the initial conditions $y(0) = y_0$ and $dy/dt(0) = y'_0$. The differentiation of Eq. (7.53) provides for the derivative of $y$

$$y'(t) = c_1 r_1 e^{r_1 t} + c_2 r_2 e^{r_2 t}. \tag{7.54}$$

The consideration of the last two formulas at $t = 0$ then provides the conditions

$$y_0 = c_1 + c_2. \tag{7.55a}$$

$$y'_0 = c_1 r_1 + c_2 r_2. \tag{7.55b}$$

We multiply Eq. (7.55a), first, with $-r_2$ and, second, with $-r_1$ and take the sum of both equations. This results in the relations

$$c_1 = \frac{y'_0 - r_2\, y_0}{r_1 - r_2}. \tag{7.56a}$$

$$c_2 = -\frac{y'_0 - r_1\, y_0}{r_1 - r_2}. \tag{7.56b}$$

With these expressions, the general solution (7.53) can be written

$$y(t) = \frac{y'_0 - r_2\, y_0}{r_1 - r_2} e^{r_1 t} - \frac{y'_0 - r_1\, y_0}{r_1 - r_2} e^{r_2 t}. \tag{7.57}$$

The validity of this solution can be proven by showing that this solution satisfies both the initial conditions (7.46) and the differential equation (7.45). This result agrees with the conclusion (5.140) obtained by the solution of the corresponding difference equation.

**Solution Features.** The features of this solution will depend essentially on $r_1$ and $r_2$. According to Eq. (7.50), we have to consider the three cases that we have two real eigenvalues ($b^2 - 4\,a\,c > 0$), one real eigenvalue ($b^2 - 4\,a\,c = 0$), and two imaginary eigenvalues ($b^2 - 4\,a\,c < 0$). Let us have a closer look at these cases to show the existence of real-valued solutions (the approach is very similar to the discussion of the corresponding discrete equation in Sect. 5.3). The first case is given for $b^2 - 4\,a\,c > 0$. The solution of the differential equation (7.45) is then given by Eq. (7.57), which can be written

$$y(t) = \frac{e^{r_1 t} - e^{r_2 t}}{r_1 - r_2} y'_0 - \frac{r_2\, e^{r_1 t} - r_1\, e^{r_2 t}}{r_1 - r_2} y_0. \tag{7.58}$$

Another convenient writing of the latter formula, which will be used for the discussion of the second and third cases below, applies the relations $r_1 = r_S + r_D$ and $r_2 = r_S - r_D$ for the eigenvalues $r_1$ and $r_2$,

$$y(t) = \frac{e^{r_D t} - e^{-r_D t}}{2 r_D} e^{r_S t}\, y'_0 - \frac{r_2\, e^{r_D t} - r_1\, e^{-r_D t}}{2 r_D} e^{r_S t}\, y_0. \tag{7.59}$$

The functions $\exp(r_D\, t)$ and $\exp(-r_D\, t)$ can be rewritten in terms of the definitions of hyperbolic sine and cosine functions (Abramowitz & Stegun 1984),

$$e^{r_D t} = \cosh(r_D t) + \sinh(r_D t), \qquad e^{-r_D t} = \cosh(r_D t) - \sinh(r_D t). \tag{7.60}$$

By using these expressions we can write the solution $y(t)$ as

$$y(t) = \frac{\sinh(r_D t)}{r_D} e^{r_S t}\, y'_0 - \frac{(r_2 - r_1)\cosh(r_D t) + (r_2 + r_1)\sinh(r_D t)}{2 r_D} e^{r_S t}\, y_0. \tag{7.61}$$

The use of $r_1 - r_2 = 2\, r_D$ and $r_1 + r_2 = 2\, r_S$ enables the following simplification of this equation,

$$y(t) = \frac{\sinh(r_D t)}{r_D} e^{r_S t}\, y'_0 + \left[ \cosh(r_D t) - \frac{\sinh(r_D t)}{r_D} r_S \right] e^{r_S t}\, y_0. \tag{7.62}$$

The second case is given for $b^2 - 4\, a\, c = 0$, which means that $r_D = 0$. The limits of hyperbolic sine and cosine functions for the case $r_D \to 0$ are given by the relations (Abramowitz & Stegun 1984)

$$\lim_{r_D \to 0} \frac{\sinh(r_D t)}{r_D} = t, \qquad \lim_{r_D \to 0} \cosh(r_D t) = 1. \tag{7.63}$$

Hence, Eq. (7.62) reads for this second case

$$y(t) = t\, e^{r_S t}\, y'_0 + (1 - r_S t)\, e^{r_S t}\, y_0. \tag{7.64}$$

The third case is given for $b^2 - 4\, a\, c < 0$, which means that $r_D$ is imaginary. To deal with this case we write

$$r_D = i\, r_{D^*}. \tag{7.65}$$

Here, $r_{D^*}$ is a real number given by

$$r_{D^*} = \frac{1}{2a} \sqrt{-b^2 + 4ac}. \tag{7.66}$$

The hyperbolic sine and cosine functions involved in Eq. (7.62) can be written for this case (Abramowitz & Stegun 1984)

$$\frac{\sinh(i\, r_{D^*} t)}{i\, r_{D^*}} = \frac{\sin(r_{D^*} t)}{r_{D^*}}, \qquad \cosh(i\, r_{D^*} t) = \cos(r_{D^*} t). \tag{7.67}$$

Hence, Eq. (7.62) reads for this third case

$$y(t) = \frac{\sin(r_{D^*} t)}{r_{D^*}} e^{r_S t}\, y'_0 + \left[ \cos(r_{D^*} t) - \frac{\sin(r_{D^*} t)}{r_{D^*}} r_S \right] e^{r_S t}\, y_0. \tag{7.68}$$

The formulas obtained for the three cases can be summarized by

$$y(t) = e^{r_S t} \begin{cases} \dfrac{\sinh(r_D t)}{r_D} y'_0 + \left[ \cosh(r_D t) - \dfrac{\sinh(r_D t)}{r_D} r_S \right] y_0 & \text{if } b^2 - 4ac > 0 \\ t\, y'_0 + (1 - r_S t)\, y_0 & \text{if } b^2 - 4ac = 0 \\ \dfrac{\sin(r_{D^*} t)}{r_{D^*}} y'_0 + \left[ \cos(r_{D^*} t) - \dfrac{\sin(r_{D^*} t)}{r_{D^*}} r_S \right] y_0 & \text{if } b^2 - 4ac < 0 \end{cases} \tag{7.69}$$

Hence, we obtain real-valued solutions for the three possible cases.

### 7.3.3 The Damped Harmonic Oscillator

**Damped Harmonic Oscillator Solution.** A specific case of the general linear second-order equation (7.45) is given by the damped harmonic oscillator equation

$$m\frac{d^2y}{dt^2}+\gamma\frac{dy}{dt}+k\,y=0. \tag{7.70}$$

The solution for this equation can be obtained by setting $a = m$, $b = \gamma$, and $c = k$ in Eq. (7.69),

$$y=e^{r_S t}\begin{cases}\dfrac{\sinh(r_D t)}{r_D}\,y'_0+\left[\cosh(r_D t)-\dfrac{\sinh(r_D t)}{r_D}r_S\right]y_0 & \text{if } \gamma^2-4mk>0\\ t\,y'_0+(1-r_S\,t)\,y_0 & \text{if } \gamma^2-4mk=0.\\ \dfrac{\sin(r_{D^*} t)}{r_{D^*}}\,y'_0+\left[\cos(r_{D^*} t)-\dfrac{\sin(r_{D^*} t)}{r_{D^*}}r_S\right]y_0 & \text{if } \gamma^2-4mk<0\end{cases} \tag{7.71}$$

According to Eqs. (7.52) and (7.66), the parameters $r_S$, $r_D$, and $r_{D^*}$ are given by

$$r_S=-\frac{\gamma}{2m},\qquad r_D=\sqrt{\left(\frac{\gamma}{2m}\right)^2-\frac{k}{m}},\qquad r_{D^*}=\sqrt{\frac{k}{m}-\left(\frac{\gamma}{2m}\right)^2}. \tag{7.72}$$

**Oscillatory Solutions.** Let us discuss the solutions of the damped harmonic oscillator equation for the case of oscillations, i.e., for $\gamma^2 - 4\,m\,k < 0$. According to Eq. (7.71), the solution for this case reads

$$\frac{y}{y_0}=e^{r_S t}\left\{\cos(r_{D^*} t)+\frac{y'_0-r_S y_0}{r_{D^*} y_0}\sin(r_{D^*} t)\right\}. \tag{7.73}$$

To simplify the discussion of this case, it is helpful to introduce an angle $\delta$ by

$$\tan\delta=\frac{y'_0-r_S y_0}{r_{D^*} y_0}. \tag{7.74}$$

For given $y_0$ and $y'_0$ values, the angle $\delta$ is determined by the expression

$$\delta=\arctan\left(\frac{y'_0-r_S y_0}{r_{D^*} y_0}\right). \tag{7.75}$$

The advantage of introducing $\delta$ is given by the possibility to rewrite the solution (7.73) in the following way (Abramowitz & Stegun 1984)

$$\begin{aligned}\frac{y}{y_0}&=e^{r_S t}\left(\cos(r_{D^*} t)+\tan\delta\sin(r_{D^*} t)\right)\\ &=e^{r_S t}\,\frac{\cos(r_{D^*} t)\cos\delta+\sin(r_{D^*} t)\sin\delta}{\cos\delta}=e^{r_S t}\,\frac{\cos(r_{D^*} t-\delta)}{\cos\delta}.\end{aligned} \tag{7.76}$$

**Nondimensional Oscillatory Solutions.** To see the damping effect on $y$ in a better way we introduce the frequency $\omega_0 = (k/m)^{1/2}$ of undamped motion and the damping frequency $\omega_d = \gamma/(2m)$. In addition, we introduce the dimensionless time $t_* = \omega_0 t$ and the dimensionless ratio $r_d = \omega_d/\omega_0$. By adopting these variables, $r_S t$ and $r_{D^*} t$ can be written

$$r_S t = -\frac{\gamma}{2m} t = -\frac{\omega_d}{\omega_0} \omega_0 t = -r_d t_*, \tag{7.77a}$$

$$r_{D^*} t = \sqrt{\frac{k}{m} - \left(\frac{\gamma}{2m}\right)^2}\, t = \sqrt{1 - \frac{m}{k}\left(\frac{\gamma}{2m}\right)^2}\, \sqrt{\frac{k}{m}}\, t = \sqrt{1 - \frac{\omega_d{}^2}{\omega_0{}^2}}\, \omega_0 t = \sqrt{1 - r_d{}^2}\, t_*. \tag{7.77b}$$

The solution (7.76) can be written then

$$\frac{y}{y_0} = \frac{e^{-r_d t_*}}{\cos\delta} \cos\left(\sqrt{1 - r_d{}^2}\, t_* - \delta\right). \tag{7.78}$$

It is convenient to introduce the nondimensional initial velocity $\lambda = y'_0/(y_0\, \omega_0)$ so that the angle $\delta$ can be written

$$\delta = \arctan\left(\frac{y'_0/(y_0\, \omega_0) - r_S/\omega_0}{r_{D^*}/\omega_0}\right) = \arctan\left(\frac{\lambda + r_d}{\sqrt{1 - r_d{}^2}}\right). \tag{7.79}$$

This writing reflects the three cases in Eq. (7.71): for $r_d = \omega_d/\omega_0 = (> 1, = 1, < 1)$ we have $\gamma^2 - 4\, m\, k = (> 0, = 0, < 0)$, respectively. The case $r_d = 1$ is referred to as critical damping, and the case $r_d > 1$ is called overdamping. For the cases $r_d = 1$ and $r_d > 1$, the use of Eq. (7.71) is the best way to find nondimensional solutions in correspondence to Eq. (7.79).

**Damping Effect.** Equation (7.78) can be used to see the damping effect, which is reflected by $r_d$. We can use Eq. (7.78) for the following observations. The most important damping effect is that the amplitude of oscillations decreases in time. A second effect is that the effective frequency becomes smaller (i.e., the period of oscillations becomes larger) due to the appearance of the square root in the cosine function. A third effect is the modification of the angle $\delta$. Figure 7.4 illustrates the damping effects. Here, $\lambda = 0$ is assumed for simplicity such that the solution $y/y_0$ does only depend on $r_d$. The amplitude reduction for $0 < r_d < 1$ may be clearly seen. The fact that damping increases the period of oscillations is clearly visible for the $r_d = 0.5$ case. For $r_d = 1$, the solution implied by Eq. (7.71) reads

$$\frac{y}{y_0} = e^{r_S t}\left(t\frac{y'_0}{y_0} + 1 - r_S\, t\right) = e^{-r_d t_*}\left(1 + (\lambda + r_d)\, t_*\right) = e^{-t_*}\left(1 + (1 + \lambda)\, t_*\right). \tag{7.80}$$

For this case, $\exp(-r_d\, t_*)/\cos\delta$ cannot be shown: we have $\delta = \pi/2$ so that $\cos\delta = 0$.

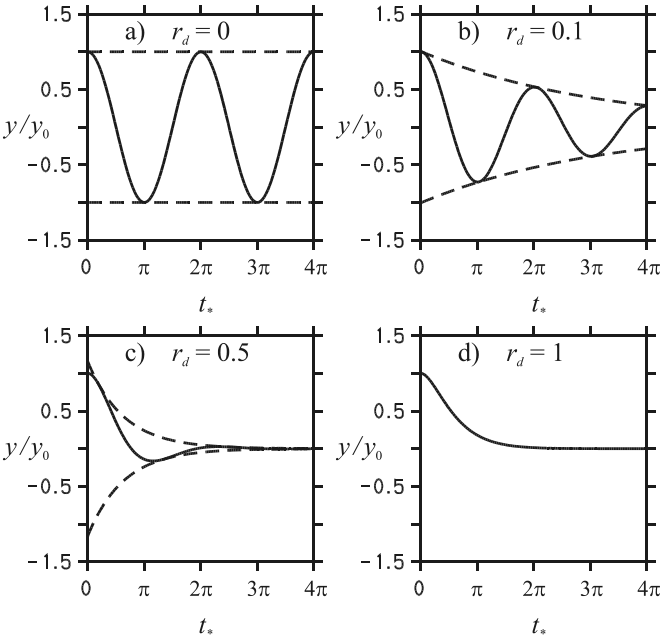



**Fig. 7.4.** The damped harmonic oscillator. The *solid lines* represent the solutions (7.78) for the $r_d$ damping values given in the figures. The *dashed lines* show the amplitude $\exp(-r_d\, t_*) / \cos\delta$.

## 7.4 Population Ecology: Growth and Self-Limitation

Let us switch now to the modeling of processes in population ecology. We will consider the evolution of a certain population in time. A population is seen as a group of individuals of the same species that live together in an area of sufficient size. The population will be measured in terms of the population density $P$, which is the number of individuals per area considered. All individuals in the population are considered to be identical, this means we do not account for age, sex, size, or other factors. For simplicity, we do not consider any immigration or emigration (which would modify the equations considered below by source and sink rates). We also do not consider any delayed responses that may lead to oscillations (such scenarios will be considered in Sect. 7.5). We also do not consider variations of the population density $P$ in space. The treatment of the latter case would require the use of more advanced mathematical equations (partial differential equations).

### *7.4.1 Growth and Self-Limitation*

Population dynamics can be affected by a lot of factors. Thus, the modeling of the evolution of populations is usually much more challenging than the modeling of mechanical processes considered before. Let us consider first some typical scenarios to illustrate the main questions. The focus here is on the representation of the modeling approach, which means the question of whether or not it is possible to postulate laws for population ecology.

**Exponential Growth.** A first case is given if a population lives under optimal conditions, this means there is no limitation of food. For this case, the population dynamics are controlled by birth and death rates, which will be proportional to the population density $P$. The evolution of $P$ in time $t$ can be described by the conservation equation

$$\frac{dP}{dt} = b\,P - d\,P = (b-d)\,P = r\,P. \tag{7.81}$$

Here, $b$ and $d$ are constant parameters that specify the birth rate $b\,P$ and death rate $d\,P$, respectively. The difference $r = b - d$ is called the rate of growth or decline, depending on whether $r$ is positive or negative. The solution to this equation can be derived by adopting the solution of the differential equation (7.7). The latter requires the replacement of $y$ by $P$, $-1/T$ by $r$, and the setting $y_e = 0$. According to Eq. (7.14), the solution of Eq. (7.81) is given by the expression

$$P = P_0\,e^{rt}, \tag{7.82}$$

where $P_0$ is written instead of $y_0$. The suitability of this solution can be seen by proving first that the initial condition is correctly recovered: by setting $t = 0$ we see that $P(0) = P_0$ as required. Next, we differentiate this solution to check that this function satisfies the differential equation (7.81),

$$\frac{dP}{dt} = r\,P_0\,e^{rt} = r\,P. \tag{7.83}$$

The last expression follows from the use of the solution (7.82) to replace the exponential function. The latter equation corresponds to the differential equation (7.81), which means that the function (7.82) is indeed the solution of Eq. (7.81). The population model (7.82) is named after the Reverend Thomas Malthus, who authored "An Essay on the Principle of Population" (1798), one of the earliest and most influential books on population dynamics. In correspondence to Newton's Laws of Mechanics, the Exponential Law (7.82), which is called the Malthusian Law, is regarded in the field of population ecology as the first principle of population dynamics.

**Self-Limitation.** The Malthusian Law model is only applicable to the case that there is no limitation of food. Obviously, this assumption is incorrect in general, this means the development of the population density in time may be limited by a maximal value. There is, therefore, the condition

$$\lim_{t\to\infty} P(t) = K. \tag{7.84}$$

Here, the constant $K$ is called the carrying capacity. This parameter determines the maximum population for which a sufficient amount of food is available in a given area (for example, the maximum number of buffalo for which grass and water was available in Northern America before 1800). The fact that the population density approaches the constant $K$ is referred to as self-limitation: the population growth leads to a self-limitation of the population because of finite resources. How can we model the population dynamics so that $P(t)$ approaches $K$ asymptotically? To have a higher flexibility (and to prepare applications to the modeling of the world population in Sect. 7.4.3), we will introduce in addition to the constant upper bound $K$ a constant lower bound $L = P(-\infty)$ that refers to the population density $P$ for $t \to -\infty$. The function $P - L$ will continuously increase from zero to $K - L$. The behavior of $P - L$ can be described, therefore, by the relation

$$P - L = (K - L)\, G(t). \tag{7.85}$$

The unknown function $G(t)$ refers to a distribution function (see Chap. 4), i.e., $G(t)$ is a non-negative function that increases monotonically from zero to one,

$$0 \le G(t) \le G(t + \Delta t) \le 1, \tag{7.86}$$

where $\Delta t$ is any non-negative time interval. The differentiation of Eq. (7.85) by $t$ provides then for $P$ the differential equation

$$\frac{dP}{dt} = (K - L)\, g(t). \tag{7.87}$$

Here, $g(t) = dG(t) / dt$ represents the population density function (PODF) related to the distribution function $G(t)$. We may assume that $g(t)$ has the properties of a probability density function (PDF). The problem to be solved now is to provide a model for $g(t)$. It is obvious that $g(t)$ can be modeled in many different ways.

## *7.4.2 Population Density Function Models*

Several possibilities to provide models for $g(t)$ will be presented in this section. The approach applied is to consider two self-limitation models. The distribution function $G(t)$ and PODF $g(t)$ that are implied by these population models will be derived as a consequence of these population models.

**Logistic Model.** First, let us consider the logistic growth model introduced by Verhulst (1838). This population model is described and used for applications in several texts (see, e.g., Edelstein-Keshet 2005, Allen 2007, Murray 2002, Boyce & DiPrima 2009, and Turchin 2003). The logistic model reads

$$\frac{dP}{dt} = \frac{1}{\tau}\frac{K-P}{K-L}(P-L) = \frac{1}{\tau}\left(1-\frac{P-L}{K-L}\right)(P-L). \tag{7.88}$$

Here, $\tau$ is constant time scale. $L = P(-\infty)$ refers to any constant initial value, and $K = P(\infty)$ refers to the constant carrying capacity. The term $P - L$ extends $P$ in the exponential model (7.81) by the consideration of a nonzero $L$, and the expression $(K - P) / [(K - L)\,\tau]$ corresponds to the growth rate $r$ in the exponential model. The idea of using the latter growth rate is to account for limitations of food due to an increasing competition for food. For small $P << K$, the growth rate is constant such that the population increases exponentially. For the case that $P$ approaches $K$ we find that the growth rate becomes zero, which means that the population levels off. The maximum rate of change $dP / dt$ can be found by considering the right-hand side of Eq. (7.88) as a function of $P$,

$$f(P) = \frac{1}{\tau}\frac{K-P}{K-L}(P-L). \tag{7.89}$$

By taking the first two derivatives we find

$$\frac{df}{dP} = \frac{1}{\tau}\frac{K-P-(P-L)}{K-L} = \frac{1}{\tau}\frac{K+L-2P}{K-L}, \tag{7.90a}$$

$$\frac{d^2 f}{dP^2} = -\frac{1}{\tau}\frac{2}{K-L}. \tag{7.90b}$$

Therefore, $dP / dt$ has a maximum at $P = (K + L) / 2$. The reason why Verhulst called his population model the logistic growth model is not fully clear. The term logistics means the delivery of required goods, at a required place, at a required time, to the required person. Possibly, Verhulst's idea of calling this equation the logistic population model was that this equation accounts for logistics issues (the requirement for food).

**Logistic Model Solution.** How can we determine the solution of the logistic model? We apply again the method of separation of variables, which requires that all variables $P$ appear on one side, and all variables $t$ appear on the other side,

$$\int \frac{(K-L)\,dP}{(K-P)(P-L)} = \int \frac{dt}{\tau}. \tag{7.91}$$

To enable the integration of the left-hand side of this relation, we have to use partial fraction expansion, this means we consider the term $(K - P)\,(P - L)$ as the

common denominator of two contributions,

$$\frac{1}{(K-P)(P-L)} = \frac{A}{K-P} + \frac{B}{P-L} = \frac{A(P-L)+B(K-P)}{(K-P)(P-L)}. \tag{7.92}$$

Here, $A$ and $B$ are any unknown constants, and the last expression results from considering the common denominator of the previous expression. The first and the last expression are consistent if $A = B = 1/(K-L)$, which means that

$$\frac{K-L}{(K-P)(P-L)} = \frac{1}{K-P} + \frac{1}{P-L}. \tag{7.93}$$

The use of this relation in Eq. (7.91) leads to

$$\int\left(\frac{1}{K-P} + \frac{1}{P-L}\right)dP = \int\frac{dt}{\tau}. \tag{7.94}$$

This equation can be integrated,

$$\ln|P-L| - \ln|P-K| = \ln\left|\frac{P-L}{P-K}\right| = \frac{t}{\tau} + C, \tag{7.95}$$

where $C$ is any constant of integration. The latter expression can be written

$$\left|\frac{P-L}{P-K}\right| = e^{C}\, e^{t/\tau}. \tag{7.96}$$

To get rid of the absolute value we write

$$\frac{P-L}{P-K} = \pm e^{C}\, e^{t/\tau} = c\, e^{t/\tau}, \tag{7.97}$$

where $\pm e^{C}$ was replaced by another unknown constant $c$. The solution reads then

$$P = \frac{L - K\,c\,e^{t/\tau}}{1 - c\,e^{t/\tau}} = \frac{K - L\,e^{-t/\tau}/c}{1 - e^{-t/\tau}/c} = \frac{K + L\,(1 - e^{-t/\tau}/c - 1)}{1 - e^{-t/\tau}/c} = L + \frac{K-L}{1 - e^{-t/\tau}/c}. \tag{7.98}$$

To determine $c$ we consider this relation at any $t_0$,

$$1 - e^{-t_0/\tau}/c = \frac{K-L}{P_0 - L}, \tag{7.99}$$

where $P_0 = P(t_0)$. Thus, $c^{-1}$ is given by

$$c^{-1} = \left(1 - \frac{K-L}{P_0 - L}\right)e^{t_0/\tau}. \tag{7.100}$$

Therefore, the solution (7.98) reads

$$P = L + (K-L)\left[1 - \left(1 - \frac{K-L}{P_0 - L}\right)e^{-(t-t_0)/\tau}\right]^{-1}. \tag{7.101}$$

The latter solution can also be written

$$\begin{aligned} P &= \frac{K+L}{2} - \frac{K-L}{2} + (K-L)\left[1 + \frac{K-P_0}{P_0-L}\, e^{-(t-t_0)/\tau}\right]^{-1} \\ &= \frac{K+L}{2} + \frac{K-L}{2}\left(2\left[1 + \frac{K-P_0}{P_0-L}\, e^{-(t-t_0)/\tau}\right]^{-1} - 1\right) \\ &= \frac{K+L}{2} + \frac{K-L}{2}\,\frac{1 - \dfrac{K-P_0}{P_0-L}\, e^{-(t-t_0)/\tau}}{1 + \dfrac{K-P_0}{P_0-L}\, e^{-(t-t_0)/\tau}}. \end{aligned} \tag{7.102}$$

**Centered Logistic Function.** A problem related to the use of the logistic function (7.102) is the need to provide the asymptotic value $K$, which is usually unknown. Instead of providing $K$, it is often more appropriate to use knowledge about the point $P_c = P(t_c)$, where $dP/dt$ has a maximum (see e.g., the discussion in Sect. 7.4.3). The discussion of the logistic model (7.88) resulted in the relation $P_c = (K + L)/2$, this means the setting of $P_c$ determines $K$. To use the assumed knowledge of $P_c$, we replace $t_0$ by $t_c$ and $P_0$ by $P_c = (K + L)/2$ in Eq. (7.102) – we can use any reference point $P_0$. By using the definition of $P_c$ we find for the ratios involved in Eq. (7.102) the expressions

$$\frac{K+L}{2} = P_c, \qquad \frac{K-L}{2} = \frac{K+L}{2} - L = P_c - L, \qquad \frac{K-P_c}{P_c-L} = \frac{K-(K+L)/2}{(K+L)/2-L} = 1. \tag{7.103}$$

The combination of the solution (7.102) with the latter expressions leads to

$$\begin{aligned} P &= P_c + (P_c - L)\,\frac{1 - e^{-(t-t_c)/\tau}}{1 + e^{-(t-t_c)/\tau}} = P_c + (P_c - L)\,\frac{e^{(t-t_c)/(2\tau)} - e^{-(t-t_c)/(2\tau)}}{e^{(t-t_c)/(2\tau)} + e^{-(t-t_c)/(2\tau)}} \\ &= P_c + (P_c - L)\tanh\left(\frac{t-t_c}{2\tau}\right) = P_c\left(1 + \frac{P_c - L}{P_c}\tanh\left(\frac{t-t_c}{2\tau}\right)\right). \end{aligned} \tag{7.104}$$

The advantage of this centered logistic function is that this model can be easily applied: for given $t_c$ and $P_c$ values, it just needs the adjustment of the time scale $\tau$ to data. The derivative of $P$, which is required for the calculation of the PODF $g$, is given by

$$\frac{dP}{dt} = \frac{P_c - L}{2\tau}\,\frac{1}{\cosh^2\left(\dfrac{t-t_c}{2\tau}\right)} = \frac{K-L}{4\tau}\,\frac{1}{\cosh^2\left(\dfrac{t-t_c}{2\tau}\right)}. \tag{7.105}$$

Here, $P_c - L = (K - L)/2$ was used in the last expression.

**Population Density Function: Logistic Model.** The PODF $g(t)$ for the logistic model can be obtained by comparing Eq. (7.105) with $dP/dt = (K - L)\,g$,

$$g(t) = \frac{1}{4\tau}\frac{1}{\cosh^2\left(\dfrac{t-t_c}{2\tau}\right)} = \frac{1}{4\tau}\frac{1}{1+\sinh^2\left(\dfrac{t-t_c}{2\tau}\right)}. \tag{7.106}$$

This function corresponds to the logistic PDF (the name of the logistic PDF arises from the fact that its distribution function is the logistic function). The mean and variance of the PODF (7.106) can be found by integrating the PODF $g(t)$ multiplied with $t - t_c$ and $(t - t_c)^2$, respectively, over $t$ from negative infinity to positive infinity. This calculation reveals that the mean is equal to $t_c$, and the standard deviation is $\tau\,\pi/3^{1/2}$: see exercise 7.4.1.

**Kapitza's Model.** Kapitza (1996) suggested a slightly different version of the population model (7.104) given by replacing the hyperbolic tangent function by an arctangent function multiplied with $2/\pi$,

$$P = P_c\left(1 + \frac{P_c - L}{P_c}\frac{2}{\pi}\arctan\left(\frac{t-t_c}{2\tau}\right)\right). \tag{7.107}$$

The reason for involving the factor $2/\pi$ is to account for the asymptotic behavior of the arctangent function. This function multiplied by $2/\pi$ approaches one for large $t$ as does the hyperbolic tangent function. Thus, we get $P(\infty) = 2\,P_c - L = K$, as required according to the definition of $K$. Justification for Kapitza's model (7.107) arises from the fact that the model (7.107) is consistent with the view of population dynamics as a self-similar process (Kapitza 1996). By using the relation $P_c - L = (K - L)/2$, the change $dP/dt$ implied by Eq. (7.107) is given by

$$\frac{dP}{dt} = \frac{P_c - L}{\pi\,\tau}\frac{1}{1+\left[(t-t_c)/(2\tau)\right]^2} = \frac{K-L}{2\pi\,\tau}\frac{1}{1+\left[(t-t_c)/(2\tau)\right]^2}. \tag{7.108}$$

**Population Density Function: Kapitza's Model.** The combination of the derivative (7.108) with the definition $dP/dt = (K - L)\,g$ of the PODF $g(t)$ shows that the PODF for Kapitza's model is given by

$$g(t) = \frac{1}{2\pi\,\tau}\frac{1}{1+\left[(t-t_c)/(2\tau)\right]^2}. \tag{7.109}$$

This is the structure of a Cauchy PDF. The mean and standard deviation can be calculated as for the logistic PODF. This calculation shows that the mean of the Cauchy PODF does not exist. The variance or higher-order central moments take reference to the mean. Therefore, these moments are not defined, too. The second moment about the center $t_c$ is infinite: see exercise 7.4.2.

### 7.4.3 Application to World Population Modeling

Let us consider the modeling of the world population development to illustrate the suitability of the concepts developed in Sects. 7.4.1 and 7.4.2. Table 7.2 and Fig. 7.5 show the corresponding data according to the Decennial Censuses, U.S. Census Bureau, U.S. Dept. of Commerce (World Almanac 2010).

**Logistic Model.** The performance of the logistic population model (7.104),

$$P = P_c\left(1 + \frac{P_c - L}{P_c}\tanh\left(\frac{t - t_c}{2\tau}\right)\right), \tag{7.110}$$

is shown in Fig. 7.5. The model parameters applied are given by

$$t_c = 1999, \qquad P_c = 6, \qquad L = 1, \qquad \tau = 30.65. \tag{7.111}$$

This model performs very well. The procedure to optimize the model parameters was the following: The value $L = 1$ is in consistency with the first data value for $t = 1804$. It is not immediately obvious which $t_c$ (and related $P_c$) should be used to characterize the position of the strongest change $dP/dt$: there are three reasonable candidates for $t_c$ given by 1987, 1999, and 2009. A good approach for finding the best choice for $t_c$ is to consider for each potential $t_c$ (and related $P_c$) the least-squares error, which reads for our case

$$E^2 = \frac{1}{N}\sum_{i=1}^{N}\left(P_i - P_M(t_i)\right)^2. \tag{7.112}$$

Here, $t_i$ and $P_i$ represent the data values given in Table 7.2, $N = 9$ is the number of samples, and $P_M$ refers to the logistic model (7.110) considered. In particular, it is helpful to consider the square root of the normalized error $E_* = E/<\widetilde{P}^2>^{1/2}$. Here, $<\widetilde{P}^2>^{1/2}$ refers to the standard deviation of population data given in Table 7.2. This normalized error $E_*$ is shown in Fig. 7.6 for the three potential $t_c$ values in dependence on the time scale $\tau$. It may be seen that $E_*$ has a minimum $E_* = 0.0337$ at $\tau = 30.65$ for the case $t_c = 1999$, for which $P_c = 6$. The latter finding leads to the parameter values given in Eq. (7.111).

**Kapitza's Model.** The corresponding performance of Kapitza's model (7.107),

$$P = P_c\left(1 + \frac{P_c - L}{P_c}\frac{2}{\pi}\arctan\left(\frac{t - t_c}{2\tau}\right)\right), \tag{7.113}$$

is also shown in Fig. 7.5, where the model parameters are given by

$$t_c = 1999, \qquad P_c = 6, \qquad L = 0.2575, \qquad \tau = 20.08. \tag{7.114}$$

Figure 7.5 shows that the performance of Kapitza's population model is comparable to the performance of the logistic model, this means both models perform equally well. The parameters for Kapitza's model were found in the following way.

**Table 7.2.** The development of the world population in time $t$ from 1804–2050 according to the Decennial Censuses, U.S. Census Bureau, U.S. Dept. of Commerce (World Almanac 2010). The population $P$ is measured in $10^9$. The population values after 2009 are projections.

| $t$ | 1804 | 1927 | 1960 | 1974 | 1987 | 1999 | 2009 | 2025 | 2050 |
|---|---|---|---|---|---|---|---|---|---|
| $P$ | 1.00 | 2.00 | 3.00 | 4.00 | 5.00 | 6.00 | 6.77 | 7.95 | 9.32 |

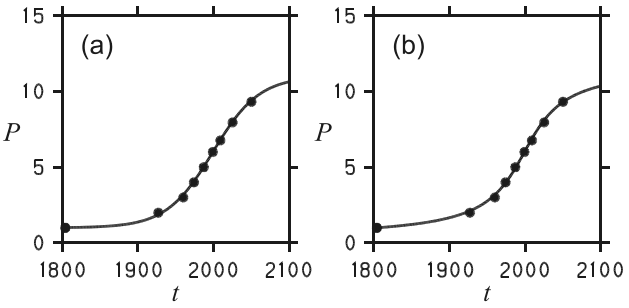



**Fig. 7.5.** The *dots* represent the world population data according to Table 7.2. (a) The *solid line* shows the prediction of the logistic model (7.110); (b) the *solid line* shows the prediction of Kapitza's model (7.113).

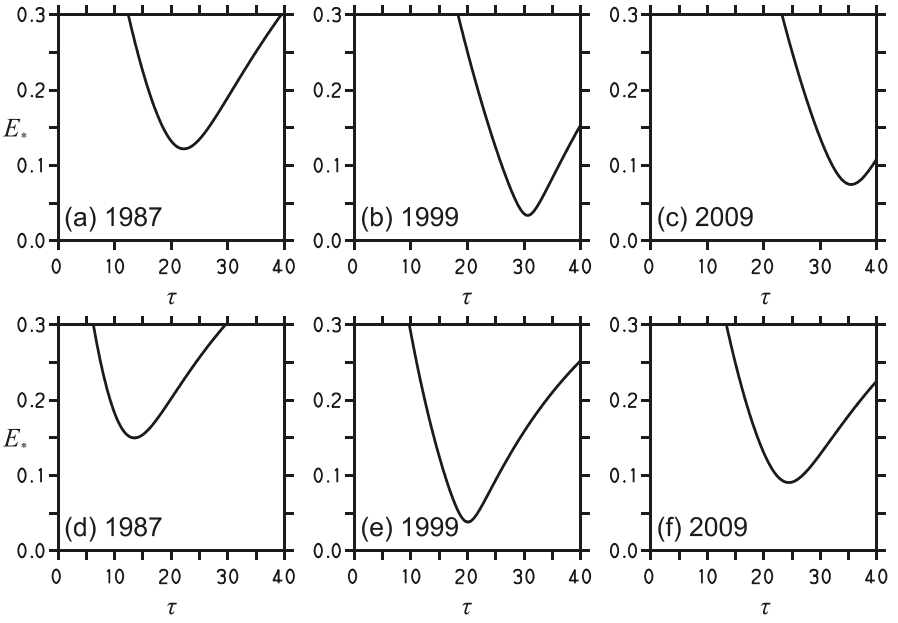



**Fig. 7.6.** The normalized least-squares error $E_* = E / <\widetilde{P}^2>^{1/2}$ for three potential $t_c$ values in dependence on the time scale $\tau$. The upper pictures apply to the logistic model (7.110). The lower pictures apply to Kapitza's model (7.113).

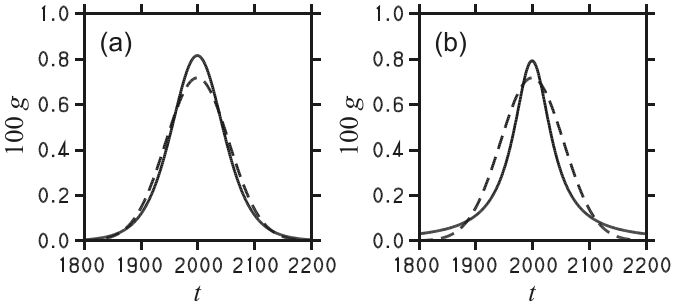



**Fig. 7.7.** PODFs $g(t)$ multiplied by 100 for two population models. The *solid lines* show in (a) the PODF (7.106) of the logistic model and in (b) the PODF (7.109) of Kapitza's model. The *dashed lines* show the normal PODF (7.117) that has the same mean and variance as the logistic PODF.

In contrast to the behavior of the logistic model, Fig. 7.5 shows that Kapitza's model provides population changes for $t < 1850$. Hence, the direct setting of any $L = P(-\infty)$ in Kapitza's model appears to be inappropriate. A way to account for this problem is to consider Kapitza's model at a data point $(t_1, P_1)$,

$$P_1 = P_c\left(1+\frac{P_c-L}{P_c}\frac{2}{\pi}\arctan\left(\frac{t_1-t_c}{2\tau}\right)\right). \tag{7.115}$$

By adopting an appropriate data point $(t_1, P_1)$ of the data given in Table 7.2, the latter relation gives the opportunity to calculate $L$,

$$L = P_c - \frac{\pi}{2}\frac{P_1-P_c}{\arctan\left(\dfrac{t_1-t_c}{2\tau}\right)}. \tag{7.116}$$

This approach was used to calculate $E_*$ depending on $\tau$ for the three potential $t_c$ values, where $(t_1, P_1) = (1804, 1)$ was used. Figure 7.6 shows that $E_*$ has a minimum $E_* = 0.0382$ at $\tau = 20.08$ for the case that $t_c = 1999$ and $P_c = 6$. The latter observation implies the parameter values given in Eq. (7.114).

**Population Density Functions.** The PODFs (7.106) and (7.109), implied by the logistic model and Kapitza's model, respectively, are shown in Fig. 7.7, where the corresponding parameters (7.111) and (7.114) are used. A PODF

$$g(t) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{(t-t_c)^2}{2\sigma^2}\right\} \tag{7.117}$$

that has the structure of a normal PDF is also shown for a comparison. Here, $\sigma = \tau\pi/3^{1/2}$ according to the standard deviation of the logistic PODF. Clearly, the logistic PODF is similar to the normal PODF: the biggest difference is the higher

peak value of the logistic PODF (which is related to the flatness $m_4 = 21 / 5 = 4.2$). On the other hand, the PODF of Kapitza's model reveals significant deviations from the normal PODF. The most important difference to the normal PODF is the very slow decay of the density function. This difference explains the fact that the Cauchy PODF implied by Kapitza's model has a nonexisting mean and central moments: the corresponding integrals cannot exist because the PODF decays too slowly for large deviations from the center $t_c$. The existence of the mean and central moments is not a required property for PDFs and PODFs, but it is clearly a desired property.

## 7.5 Population Ecology: Oscillations and Collapse

Growth and self-limitation discussed in Sect. 7.4 represent two basic features of population dynamics, but there may be also other basic processes that have to be considered. The purpose of the discussion in this section is to present a more general picture of population changes. On this basis, we will consider the question of which trend will possibly be seen regarding the future development of the world population.

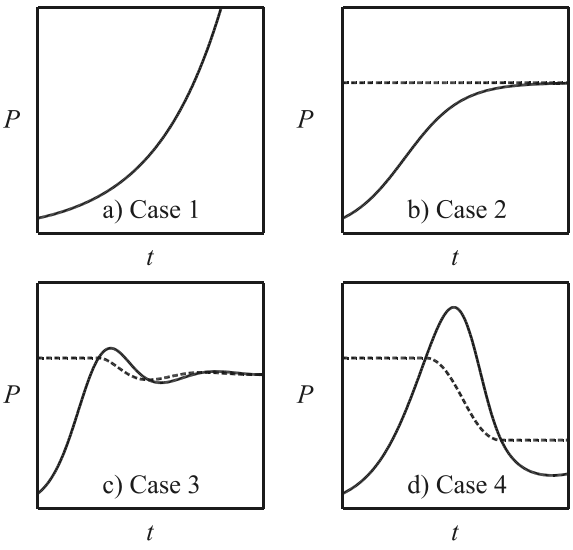



**Fig. 7.8.** Basic scenarios of population dynamics: (a) continuous growth, (b) balance, (c) oscillation, and (d) collapse. The *solid lines* present the population density $P$, and the *dashed lines* present the corresponding carrying capacity $K$ (see the explanations in Sect. 7.5.3).

### *7.5.1 Basic Population Dynamics Scenarios*

Figure 7.8 illustrates basic possibilities (Meadows et al. 2004) in which a population can interact with its asymptotic limit, the carrying capacity. Let us have a closer look at these scenarios in order to prepare the following discussions in this section.

**Case 1 and Case 2.** A case 1 illustrated in Fig. 7.8a, which was considered in Sect. 7.4, is given by a population that does continuously grow if the carrying capacity is much higher than the population values, or if the carrying capacity increases faster than the population. A case 2 illustrated in Fig. 7.8b, which was also considered in Sect. 7.4, is given by a population that levels off smoothly below the carrying capacity. Examples for these two cases were given by the exponential increase of the U.S. population from 1790 to 1890 discussed in Chap. 1 and the world population growth discussed in Sect. 7.4.

**Case 3 and Case 4.** There are also other population dynamics scenarios that may be observed if the population overshoots the carrying capacity. Overshooting may occur if there are delayed responses involved such that the population does not immediately experience the limiting action of the carrying capacity. A case 3 illustrated in Fig. 7.8c is given if the population overshoots the carrying capacity without doing massive and permanent damage to the system. In such a case, the population will oscillate about the carrying capacity and level off after establishing a new balance with the carrying capacity. Another possibility is a case 4 illustrated in Fig. 7.8d: the population does overshoot the carrying capacity with severe and permanent damage of system resources. In such a case, the population will rapidly decline to achieve a new balance with the reduced carrying capacity at a much lower level. This case will imply a collapse of the original system. The term collapse used here does not mean the extinction of a population, but it is used to refer to the fact that the original system (characterized by the original constant carrying capacity) does not exist anymore. An example for these cases is the following: deer or goats, when natural enemies are absent, often overgraze their range, which causes erosion or destruction of the vegetation (Kormondy 1996). The consequences of such overgrazing are usually seen with a typical delay time of several months.

**Future of Mankind.** An obvious question regarding the development of the world population is whether and under which conditions the world population and related factors will grow, hold steady, oscillate, or collapse in the future: which out of the four scenarios described in the preceding paragraph will be realized? The latter question will be addressed in Sects. 7.5.2 and 7.5.3 by presenting two population models and their conclusions: the World3 model and a logistic model that can account for delay effects.

### *7.5.2 The World3 Model*

**World3 Model.** In 1968, thirty individuals from ten countries (scientists, educators, economists, humanists, industrialists, and international civil servants) gathered in Rome to discuss a subject of stagerring scope: the present and future predicament of mankind. The purpose of this meeting was to examine the "world problematique": poverty in the midst of plenty, degradation of the environment, loss of faith in institutions, uncontrolled urban spread, insecurity of employment, alienation of youth, rejection of traditional values, and inflation and other monetary disruptions (Meadows et al. 1972). Out of this meeting in 1968 grew the Club of Rome, which is an informal organization. Its purposes are to foster understanding of the global economic system in order to bring that new understanding to the attention of policy-makers and the public worldwide, and to promote in this way new policy initiatives and actions. The Club of Rome commissioned a team of analysts from the Massachusetts Institute of Technology (MIT) to analyze the "world problematique" using a computer model called World3 developed at MIT. The time scale for the model began in the year 1900 and continues until 2100. Historical values to the year 1970 are broadly reproduced in the World3 output. The results of this analysis were published for the general public by Meadows et al. (1972). A detailed description of the World3 model is available in the book "Dynamics of Growth in a Finite World" by Meadows et al. (1974a). A 30-year update of the 1972-analysis was published by Meadows et al. (2004). Recently, Turner (2008) compared the World3 model output (Meadows et al. 1974b) with 30 years of reality (for 1970–2000). The following presentation in this subsection follows Turner's (2008) analysis.

**World3 Model Structure and Variables.** The World3 model involves four key elements. A first key element is the presence of resources, such as agricultural land that may be eroded as a result of the functioning of the economic system. A second key element is the presence of delays in the signals from one part of the system to another. For example, the effect of increasing pollution levels (see the discussion of global warming in Chap. 1), may not be recognized for some decades. A third key element is the consideration of both positive and negative feedback loops. The interaction of such feedback loops may lead to oscillations of variables over time. The fourth key element is given by the consideration of the global economic system as a complete system of several subsystems: population, food production, industrial production, consumption of nonrenewable natural resources, and global pollution. A detailed explanation of the meaning of these model variables is given in Turner's (2008) review "A comparison of *The Limits to Growth* with 30 years of reality".

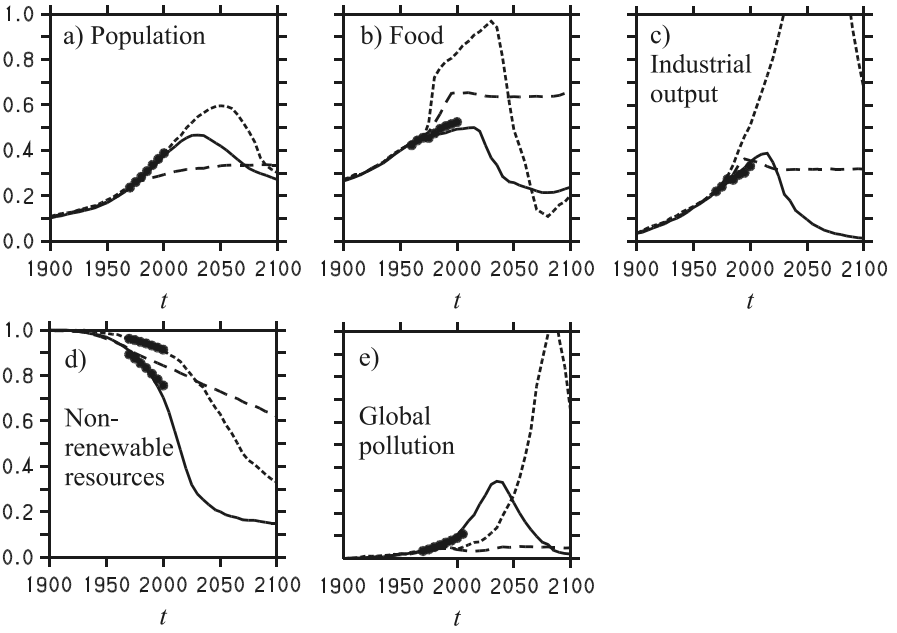



**Fig. 7.9.** The World3 results. The *solid lines* present the standard scenario World3 results, the *hatched lines* present the comprehensive technology scenario results, and the *dashed lines* present the stabilized world scenario results (Meadows et al. 1974b). Historical values to the year 1970 are reproduced by all scenarios with the exception of nonrenewable resources of the comprehensive technology scenario. For this case, the natural resource reserves were assumed to be twice as high as for the other cases. The *dots* present observed data for the period 1970–2000 according to Turner's (2008) analysis. The observed data were normalized to the World3 model values for 1970. Regarding the nonrenewable resources, the *lower dots* and *upper dots* provide minimum and maximum estimates. The global $CO_2$ concentration was used to obtain the global pollution curves.

**World3 Model Scenarios.** The analysis of Meadows et al. (1974b) considered the following three basic scenarios that are illustrated in Fig. 7.9:

- standard scenario:

  This scenario assumes a business-as-usual situation. Parameters reflecting physical, economic, and social relationships are maintained at values consistent with the period 1900–1970. This scenario results in "overshoot and collapse" of the global system (see the case 3 scenario of Fig. 7.8) about midway through the 21st century due to a combination of diminishing resources and increasing ecological damage due to pollution.

- comprehensive technology scenario:
  This approach attempts to solve sustainability issues with a broad range of purely technological solutions. The scenario considers levels of resources that are effectively unlimited. It is assumed that 75% of materials are recycled, pollution generation is reduced to 25% of its 1970 value, agricultural land yields are doubled, and birth control is available worldwide. These efforts delay the collapse (see the case 3 scenario of Fig. 7.8) of the global system to the later part of the 21st century, when the growth in economic activity has outstripped the gains in efficiency and pollution control.
- stabilized world scenario:
  Both technological solutions and deliberate social policies are implemented to achieve equilibrium states for key factors including population, material wealth, food, and services per capita (the case 2 scenario of Fig. 7.8). Such equilibrium states may be achieved, for example, by perfect birth control, a pollution control technology, and a maintenance of agricultural land. The variables in Fig. 7.9 approach equilibrium values in this case (except the nonrenewable resources).

**Model Evaluation.** Turner (2008) compared World3 predictions with 30 years of reality (which means with observed data for 1970–2000). He concluded that these observations compare favorably with key features of the standard scenario (business-as-usual) predicting "overshoot and collapse" of the global system about midway through the 21st century. The data do not compare well with the other two scenarios (the comprehensive technology and stabilized world scenarios).

### *7.5.3 A Delay Logistic Model*

**Delay Logistic Model.** The World3 model is very complex such that it cannot be easily studied. Let us cover some basic assumptions of this model on the basis of a much simpler model. In particular, let us consider the following extension of the logistic model (7.88),

$$\frac{dP}{dt} = \frac{1}{\tau}\left(1 - \frac{M}{K - L}\right)(P - L). \tag{7.118}$$

This delay logistic model involves a memory term

$$M(t) = \int_{-\infty}^{t} \mu(t - s)\big(P(s) - L\big)\, ds. \tag{7.119}$$

The function $\mu(t - s)$ is a memory function: all the population history is used for the calculation of the change $dP/dt$ at $t$. The memory effect also represents a delay effect: the effect of the population $P(s)$ with $-\infty < s \le t$ on $dP/dt(t)$ can be seen as a delayed effect. The model (7.118) is unclosed as long as the memory function

$\mu(t-s)$ is not defined. The carrying capacity will be considered as a function of $t$ here, $K = K(t)$, to allow dynamic interactions between the population density $P$ and $K$. The model parameter $L$ represents again any constant initial population level, $L = P(-\infty)$, and the characteristic time scale $\tau$ is a constant, too.

**Definition Memory Function.** The memory function can be defined in many different ways (Allen 2007). Here, we consider the model

$$\mu(t-s) = \frac{1}{\tau_M}\exp\left\{-\frac{|\,t-s\,|}{\tau_M}\right\}, \tag{7.120}$$

where $\tau_M$ refers to a characteristic memory time scale. This memory functions has the normalization property to integrate to one,

$$\int_{-\infty}^{t}\mu(t-s)\,ds = \frac{1}{\tau_M}\int_{-\infty}^{t}\exp\left\{-\frac{t-s}{\tau_M}\right\}ds = \exp\left\{-\frac{t-s}{\tau_M}\right\}_{s=t} - \exp\left\{-\frac{t-s}{\tau_M}\right\}_{s=-\infty} = 1. \tag{7.121}$$

How are the logistic model (7.88) and the delay logistic model (7.118) combined with Eq. (7.120) related to each other? The consideration of the memory term in the limit $\tau_M \to 0$ shows the following,

$$\begin{aligned}\lim_{\tau_M\to 0} M(t) &= \int_{-\infty}^{t}\big(P(s)-L\big)\lim_{\tau_M\to 0}\mu(t-s)\,ds \\ &= \big(P(t)-L\big)\int_{-\infty}^{t}\lim_{\tau_M\to 0}\mu(t-s)\,ds = P(t)-L.\end{aligned} \tag{7.122}$$

The conclusion that $P(s) - L$ can be replaced by $P(t) - L$ can be justified in terms of the properties of the memory function: $\mu(t-s)$ is always zero for $\tau_M \to 0$ except at $s = t$ (i.e., $\mu(t-s)$ behaves like a delta function). Therefore, only $P(t) - L$ will contribute to the integral. The last expression in Eq. (7.122) arises from the normalization property (7.121), which applies to all $\tau_M$. Thus, we can conclude that the delay logistic model (7.118) combined with Eq. (7.120) recovers the logistic model (7.88) in the limit $\tau_M \to 0$.

**Model Reformulation.** The analysis and application of the model (7.118) are complicated because of the memory term $M(t)$. However, the use of the properties of the memory function $\mu(t-s)$ enables a much simpler representation of the delay logistic model. First, we differentiate $M(t)$ to obtain for $M$ the equation

$$\begin{aligned}\frac{dM}{dt} &= \int_{-\infty}^{t}\frac{d\mu(t-s)}{dt}\big(P(s)-L\big)\,ds + \mu(0)\big(P(t)-L\big) \\ &= -\frac{1}{\tau_M}\int_{-\infty}^{t}\mu(t-s)\big(P(s)-L\big)\,ds + \frac{P(t)-L}{\tau_M} = \frac{P(t)-L-M(t)}{\tau_M},\end{aligned} \tag{7.123}$$

**Table 7.3** Parameters of the model (7.124) for the curves shown in Fig. 7.10.

| | $L$ | $P_0$ | $M_0$ | $K_0$ | $\tau$ | $\tau_M$ | $\tau_K$ |
|---|---|---|---|---|---|---|---|
| Case 1 | 0 | 0.1 | – | $\infty$ | 60 | – | – |
| Case 2 | 0 | 0.1 | 0.1 | 1 | 25 | 0 | – |
| Case 3 | 0 | 0.1 | 0.1 | 1 | 15 | 10 | 20 |
| Case 4 | 0 | 0.1 | 0.1 | 1 | 25 | 40 | 40 |

where the definitions of $M$ and $\mu(t-s)$ are applied. By making use of this equation for $M$ we can write the model (7.118) in terms of an equation system for $P$ and $M$,

$$\frac{dP}{dt} = \frac{1}{\tau}\left(1 - \frac{M}{K-L}\right)(P-L), \tag{7.124a}$$

$$\frac{dM}{dt} = \frac{P-L-M}{\tau_M}. \tag{7.124b}$$

Equation (7.124b) enables the derivation of the conclusion (7.122) for $\tau_M \to 0$ in a simple way: by multiplying this equation with $\tau_M$ and using $\tau_M \to 0$ we recover $M = P - L$. The equation system (7.124) has to be combined with appropriate initial values at any time $t = 0$,

$$P_0 = P(0), \tag{7.125a}$$

$$M_0 = M(0) = \int_{-\infty}^{0} \mu(-s)\big(P(s) - L\big)\, ds. \tag{7.125b}$$

Relation (7.125b) shows that the previous history of $P(s)$ with $-\infty < s \le 0$ has to be provided to calculate $M_0$. We will assume here that $P(s) = P_0$ for $-\infty < s \le 0$. For this case, $M_0$ is given by

$$M_0 = \big(P_0 - L\big)\int_{-\infty}^{0} \mu(-s)\, ds = P_0 - L. \tag{7.126}$$

The last expression follows from Eq. (7.121) by setting $t = 0$. Unfortunately, it is impossible to solve the equation system (7.124) analytically, i.e., these equations have to be solved numerically.

**Four Scenarios.** The population model (7.124) is capable of covering the four basic scenarios of population dynamics shown in Fig. 7.8. This will be made clear in the following by using four definitions (corresponding to the four scenarios considered) for the carrying capacity $K(t)$. The four modeling approaches applied for $K(t)$ will be explained in terms of Fig. 7.10. This figure agrees with Fig. 7.8: the difference is that the model parameter values (which are given in Table 7.3) and function values are specified in Fig. 7.10.

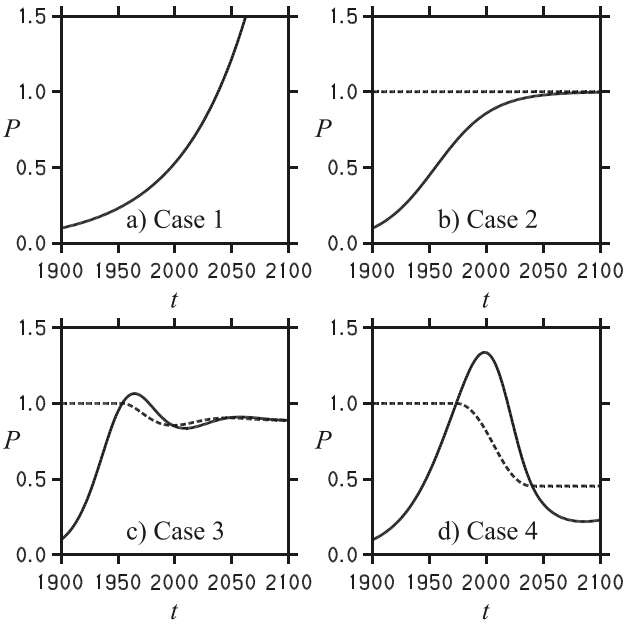



**Fig. 7.10.** Basic scenarios of population dynamics: (a) case 1: continuous growth, (b) case 2: balance, (c) case 3: oscillation, and (d) case 4: collapse. The *solid lines* show the population density $P$, and the *dashed lines* show the carrying capacity $K$. The numbers applied here are arbitrary (an initial carrying capacity $K_0 = 1$ was assumed). The model parameters are given in Table 7.3.

• Case 1: The first case assumes an infinite value of $K$,

$$K \to \infty. \tag{7.127}$$

Correspondingly, the population model (7.124a) is equivalent to the use of the exponential population growth model

$$\frac{dP}{dt} = \frac{P - L}{\tau}. \tag{7.128}$$

This model was used to obtain Fig. 7.10a that shows an exponential increase.

• Case 2: The second case assumes a constant and finite value of $K$,

$$K = K_0. \tag{7.129}$$

Here, $K_0$ refers to the initial value of $K$. This case corresponds to the consideration of a logistic model if there are no memory effects, i.e., if $\tau_M = 0$. This model was applied to obtain Fig. 7.10b (this case can be handled numerically by replacing Eq. (7.124b) for the calculation of $M$ by $M = P - L$, see the discussion below Eq. (7.124)). Another case 2M (the M refers to the inclusion of memory effects) is

given if $\tau_M$ is nonzero. The case 2M enables the simulation of oscillations about the carrying capacity $K$. Finally, a stationary state is established where

$$0 = \frac{1}{\tau}\left(1 - \frac{M}{K - L}\right)(P - L), \tag{7.130a}$$

$$0 = \frac{P - L - M}{\tau_M}. \tag{7.130b}$$

This stationary state is realized by the values $M = K - L$ and $M = P - L$, which implies that $P = K$. An interesting question is under which condition the logistic-shape behavior of curves changes such that $P$ overshoots the carrying capacity $K$. The latter question is addressed in terms of Fig. 7.11 that illustrates the effect of the memory time scale $\tau_M$. We see that $\tau_M / \tau$ has to be sufficiently large to enable an overshooting. The overshooting is observed if $\tau_M \geq \tau / 3$: in this case we find that $(P - K) / K \geq 10^{-5}$. Figure 7.11a presents $P(t)$ for the critical value $\tau_M = \tau / 3$, and Figure 7.11b shows the overshooting effect for $\tau_M = 3\ \tau / 2$. Interestingly, the critical value $\tau_M = \tau / 3$ was also found for variations of model parameters (e.g., other $\tau$, $P_0 = M_0$, and $K_0$ values than applied for the cases here).

• Case 3: The third case assumes that $K$ changes in interaction with $P$,

$$\frac{dK}{dt} = \theta(t - t_K)\frac{K - P}{\tau_K}. \tag{7.131}$$

Here, $\tau_K$ is a characteristic time scale for changes of $K$, and $t_K$ is the smallest time for which $P(t_K) = K$. The idea of this model is that $K$ begins to decrease if $P$ overshoots the carrying capacity, $P > K$. According to this model, it is possible that $K$ increases at later times if $K > P$ (which models a recovery of the carrying capacity), see Fig. 7.10c.

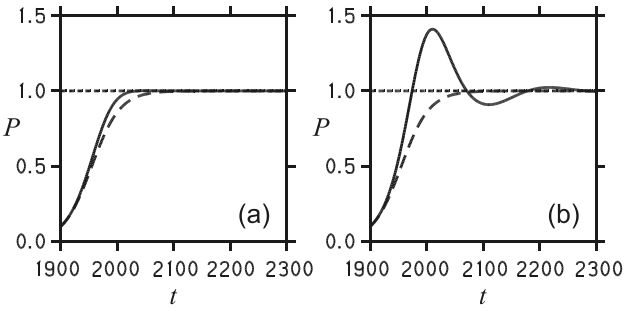



**Fig. 7.11.** The *solid lines* present the case 2M scenario, where $L = 0$, $P_0 = M_0 = 0.1$, $K_0 = 1$, and $\tau = 25$. The difference between these lines is the use of $\tau_M = \tau / 3$ and $\tau_M = 3\ \tau / 2$ in (a) and (b), respectively. The *dashed lines* show the corresponding case 2, where $\tau_M = 0$ (see Table 7.3); the *hatched lines* show $K = K_0 = 1$.

**Table 7.4** Parameters of the model (7.124) for the curves shown in Fig. 7.12.

| | $L$ | $P_0$ | $M_0$ | $K_0$ | $\tau$ | $\tau_M$ | $\tau_K$ |
|---|---|---|---|---|---|---|---|
| Case 2 | 1 | 1.0151 | 0.0151 | 11 | 30.65 | 0 | – |
| Case 2M | 1 | 1.0151 | 0.0151 | 10.52 | 30.95 | 4.40 | – |

• Case 4: The fourth case also assumes that $K$ changes in interaction with $P$,

$$\frac{dK}{dt} = \min\left(0, \frac{K-P}{\tau_K}\right) \tag{7.132}$$

Here, $K$ decreases as in the case 3 model if $K > P$, but an increase of $K$ is impossible ($K$ does not change if $K > P$), see Fig. 7.10d.

**World Population Modeling.** Let us apply the population model (7.124) to the modeling of the world population. The population data considered are given in Table 7.2. Obviously, the case 1 modeling approach of infinite growth does not apply to this data set. For the case considered we have $P \le K$. This fact suggests the use of the case 2 (for which $\tau_M = 0$) or case 2M (for which $\tau_M \neq 0$) models that apply a constant $K$, this means it is expected that $K$ does not change due to an overshooting of $P$. It was proven that the case 3 and case 4 modeling approaches do not represent an alternative to the use of the case 2 and case 2M approaches. The case 2 model parameters correspond to the data used for the logistic model (7.110) combined with Eq. (7.111). Here, $P_0 = P(t = 1800)$, where $P(t)$ is given by Eq. (7.110). The curve for the case 2 model agrees exactly with the logistic model (7.110). The parameters for the case 2M model were obtained by keeping $M$, $P_0$, and $M_0 = P_0 - L$ fixed and optimizing $K_0$, $\tau$, and $\tau_M$ such that the least-square error becomes minimal. A way to realize this nontrivial optimization problem is to optimize $K_0$ for given $\tau$ and $\tau_M$ values, and to consider then the effect of $\tau$ and $\tau_M$ variations. The normalized error $E_* = 0.0271$ that results from this optimization of parameters is smaller than the corresponding normalized error $E_* = 0.0337$ related to the logistic curve. Hence, the case 2M model is slightly more accurate than the case 2 model. Figure 7.12 shows a comparison of the case 2 model and case 2M model. The corresponding model parameters are given in Table 7.4. There is hardly any observable difference between the two models. The latter comparison confirms the World3 assumption about a nonzero memory time $\tau_M$. However, the small memory time $\tau_M = 4.4$ years obtained here means that the basic feature of the logistic curve to approach $K$ smoothly (so that $P \le K$) is unchanged. According to the discussion of the case 2M above, it would need $\tau_M > \tau/3$ to enable values $P > K$. Hence, this analysis does not support the case 4 scenario features that are obtained by the World3 model (see Figs. 7.8 and 7.9).

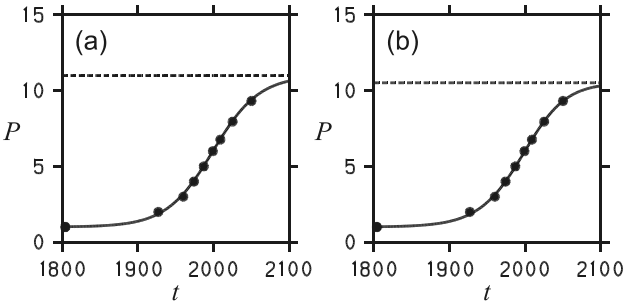



**Fig. 7.12.** World population modeling based on the model (7.124). The *solid lines* in (a) and (b) show the result of using the case 2 model and the case 2M model, respectively. The model parameters for these cases are given in Table 7.4. The *hatched lines* show the corresponding carrying capacity $K$. The *dots* present the observed data and predictions according to Table 7.2.

## 7.6 Summary

This chapter described the mathematical analysis of a variety of processes. The objective was to find equations that we can use for the analysis and prediction of processes. Actually, our ultimate goal is to find general equations (laws) that can be applied to a relatively broad range of processes. Did we accomplish this goal? Let us address the latter question by having a closer look at the properties of a law and the equations that we found for mechanical and population ecology processes.

**Laws.** The term law is often used with different meanings. Here, the term law refers to a mathematical model that represents an idea about the mechanism of a processes considered. An important property of a law is that it has to be proven to be correct: there has to be convincing evidence for this law given by observations (e.g., measurements that can be repeated). Unfortunately, it will never be possible to have only one equation for everything. The reason is simply that an equation has to explain the mechanism of processes, and there is a huge variety of different processes. Consequently, we have to use four population models for four different population processes. There may be different (mechanical or population ecology) processes that look relatively similar, e.g., a simple balance of heat transfer and the logistic population curve. However, also for such similar processes we have to use different models because the nature of these processes is different. Hence, the goal can only be to find laws that cover certain processes. A law usually involves model parameters that enable the application of the model to a certain range of processes.

**Laws of Mechanics.** We discussed various mechanical processes: the distribution of heat or mass, and oscillations. Regarding the distribution of heat we saw that there exists a theoretical framework given by Newton's Law of Cooling, and a corresponding equation can be applied to the modeling of the distribution of mass. Motions of macroscopic bodies (moving with velocities much smaller than the speed of light) can be described very well by Newton's Laws of Motion. Thus, we have laws that we can use to describe mechanical processes. This does not mean that we have completely specified equations for every possible case. For example, the suitability of using a damping force that is a linear function of the velocity requires evidence (depending on the case considered it may be more appropriate to use other damping force models). The meaning of the conclusion that there exist laws for the mechanical processes considered here is that there is an established mathematical basis for addressing modeling problems.

**Laws of Population Ecology.** We also considered the question of whether or not there are laws of population ecology. This question is more complicated than the question about laws for mechanical processes because population systems are often much more complex than mechanical systems. There are different opinions about the existence of population dynamics laws. One widespread opinion among ecologists is that ecology lacks general laws. On the other hand, Turchin (2001, 2003) argues that there are some very general law-like propositions that provide the theoretical basis for most population dynamics models. Scheiner and Willig (2005, 2008) argue that the fundamental principles (similar to Darwin's (1859) nonmathematical theory of evolution) of ecology are known, but these principles are not yet manifested in mathematical models. The conclusion that can be drawn from the discussion here is that there is currently no law of population ecology in the sense that there is one equation that applies to a variety of similar processes. However, this situation may change. A basic element that is missing in equations for population dynamics is the PODF, which has all the properties of a probability density function (PDF). By following the modeling concepts presented in Chap. 8 it is possible to postulate an equation for the PODF.

## 7.7 Exercises

**7.2.1** A cup of coffee at 90°C is poured into a mug and left in a room at 21°C. After one minute, the coffee temperature is 85°C. Suppose that the coffee temperature does obey Newton's Law of Cooling. The coffee becomes safe to drink after it cools to 60ºC. How long will it take before you can drink the coffee, this means at which time is the coffee temperature 60ºC?

**7.2.2** A cup of coffee is poured into a mug and left in a room at 15°C. After one minute, the coffee temperature is 80°C. After two minutes, the coffee temperature is 72°C.

a) Use Newton's Law of Cooling for calculating the coffee temperature at the initial time.

b) Determine the coffee temperature after five minutes.

**7.2.3** Newton's Law of Cooling assumes that the temperature decrease is proportional to the difference between the temperature and the constant external temperature. For cooling in still air, a better model is to assume that the temperature decrease is proportional to the $5/4^{\text{th}}$ power of the difference between the temperature and the constant external temperature.

a) Formulate the corresponding differential equation.

b) Calculate the solution of this differential equation by using the method of the separation of variables.

c) Assume that the initial temperature is equal to the external temperature. Use the solution derived in b) to determine the temperature for all $t$.

**7.2.4** Solve the coffee temperature problem described in exercise 7.2.1 by using the solution obtained in exercise 7.2.3.

**7.2.5** Stefan-Boltzmann's Law of Radiation states that the temperature change $dT/dt$ of a body at $T$ degrees Kelvin is proportional to $E^4 - T^4$, this means that $dT/dt = k\,(E^4 - T^4)$. Here, $E$ refers to the constant absolute temperature of the surroundings (measured in degrees Kelvin), and $k$ is a constant.

a) Rewrite the differential equation $dT/dt = k\,(E^4 - T^4)$ by introducing a nondimensional temperature $T_*$ and nondimensional time $t_*$.

b) Determine the solution $T_*$ of the nondimensional differential equation by using the method of the separation of variables. Hint: use the following integral ($a$ and $C$ refer to any constants):

$$\int \frac{dx}{a^4 - x^4} = \frac{1}{4a^3}\ln\left(\frac{a+x}{a-x}\right) + \frac{1}{2a^3}\arctan\left(\frac{x}{a}\right) + C.$$

c) Explain how the dependence of $T_*$ on $t_*$ can be graphically illustrated for a given value of the initial nondimensional temperature $T_{*0}$.

**7.2.6** A swimming pool has a volume of 50 $m^3$. A mass $C$ (in kg) of chlorine is dissolved in the pool water. Starting at a time $t = 0$, water containing a concentration of $0.1\ C/V$ chlorine is pumped into the swimming pool at a rate of 0.02 $m^3$ / min, and the water flows out at the same rate.

a) Present the differential equation for the chlorine mass $Q(t)$.

b) Find the solution $Q(t)$ to this equation.

c) What is the amount of the chlorine mass $Q(t)$ after 2 hours?
d) At which time is the chlorine mass in the pool 50% of the initial mass?

**7.2.7** Consider the following problem (Boyce & DiPrima 2009). A swimming pool has a volume of 50 $m^3$. The water has been contaminated by 5 kg of a nontoxic dye that leaves a swimmer's skin an unattractive green. The filter system can take water from the pool, remove the dye, and return the water to the pool at a flow rate of 0.02 $m^3$ / min.
a) Present the equation for the temporal development of the dye mass $Q(t)$.
b) Find the solution $Q(t)$ to this equation.
c) The effect of the dye is imperceptible if its concentration $c = Q / V$ is less than 5 g / $m^3$. How long does it take to reduce the dye concentration to this value?

**7.3.1** A projectile is fired with an angle of elevation $\alpha$ and an initial velocity given by $\boldsymbol{v}_0 = v_0 \cos\alpha\, \boldsymbol{i} + v_0 \sin\alpha\, \boldsymbol{j}$. Here, $\boldsymbol{i}$ and $\boldsymbol{j}$ refer to unit vectors in the horizontal $x$ direction and vertical $y$ direction, respectively. The projectile motion is only affected by the gravity force $\boldsymbol{F} = -\, m\, g\, \boldsymbol{j}$, where $m$ is the mass of the projectile and $g$ is the gravity acceleration. Use Newton's Second Law to do the following:
a) Calculate the velocity vector $\boldsymbol{v}(t)$.
b) Calculate the position vector $\boldsymbol{r}(t)$. The initial position vector is $\boldsymbol{r}(0) = \boldsymbol{0}$.
c) Calculate the horizontal distance $H$ traveled by the projectile.
d) Calculate the value of $\alpha$ for which the distance $H$ has a maximum.

**7.3.2** Consider the following initial value problem ($A$ is any parameter):

$$\frac{d^2y}{dt^2} + (A+4)\frac{dy}{dt} + (A+3)\,y = 0, \qquad y(0) = 1, \quad \frac{dy}{dt}(0) = 0.$$

a) Find the solution to this initial value problem.
b) For which values of the parameter $A$ are solutions bounded as $t \to \infty$?
c) Find the critical $t$ value at which $y(t)$ may have an extreme value.
d) Consider the parameter $A$ range for which we have bounded solutions. Does $y(t)$ have a local minimum or maximum at this critical $t$ value?

**7.3.3** Consider the following initial value problem ($A$ is any parameter):

$$\frac{d^2y}{dt^2} + 4\frac{dy}{dt} + 4\,y = 0, \qquad y(0) = 1, \quad \frac{dy}{dt}(0) = A.$$

a) Find the solution to this initial value problem.
b) Find the critical $t$ value at which $y(t)$ may have an extreme value.
c) For which values of the parameter $A$ do we find at the critical $t$ value a local minimum, a local maximum, and no extreme value of $y(t)$?

**7.3.4** Find a differential equation that has the general solution $y = c_1 e^{3t} + c_2 e^{-t}$.

**7.3.5** The displacement of a spring-mass system is described by the initial value problem ($y_0$ is any parameter)

$$\frac{d^2y}{dt^2} + 4\frac{dy}{dt} + 8y = 0, \qquad y(0) = y_0, \quad \frac{dy}{dt}(0) = 0.$$

a) Find the solution to this initial value problem.
b) Write the solution derived in a) so that the time dependence does only enter via an exponential function and a cosine function. Hint: follow the transition from Eq. (7.73) to Eq. (7.76).

**7.3.6** Consider the damped harmonic oscillator solution (7.76).
a) Calculate the time $\tau$ between any successive maxima.
b) Calculate the ratio $R$ of the displacements of any successive maxima.
c) Explain how the damping constant $\gamma$ can be calculated via measurements of $R$, $\tau$, and the mass $m$.

**7.3.7** Electrical vibrations in electric circuits can be described by the following differential equation for the charge $Q(t)$ measured in coulombs (C),

$$L\frac{d^2Q}{dt^2} + R\frac{dQ}{dt} + \frac{1}{C}Q = 0.$$

Here, $R$ refers to the resistance, which is measured in ohms (Ω), $L$ refers to the inductance, which is measured in henrys (H = Ω s), and C refers to the capacitance, which is measured in farads (F = s / Ω). The variables $L$, $R$, and $C$ are non-negative and assumed to be constant. An impressed voltage, which can involve a forcing on the right-hand side of the equation, is not considered. The equation considered represents Kirchhoff's Second Law.
a) Assume that $L$ = 2 H, $R$ = 20 Ω, and $C$ = 0.01 F. Find $Q(t)$ for the case that $Q(0)$ = 5 C and $dQ/dt\,(0) = 0$.
b) Assume that $L$ = 2 H and $C$ = 0.01 F. Which $R$ value is required so that the circuit is critically damped?
c) Find the charge $Q(t)$ for the case of a critical damping considered in b). Assume that $Q(0)$ = 5 C and $dQ/dt\,(0) = 0$.

**7.4.1** Consider the logistic PODF given by Eq. (7.106).
a) Calculate the mean implied by this PODF.
b) Calculate the standard deviation implied by this PODF.
c) Calculate the skewness and flatness implied by this PODF.
Hint: use the following integral values, which can be found numerically.

$$\int_{-\infty}^{\infty} \frac{x^2}{\cosh^2(x)}dx = \frac{1}{6}\pi^2, \qquad \int_{-\infty}^{\infty} \frac{x^4}{\cosh^2(x)}dx = \frac{7}{120}\pi^4.$$

**7.4.2** Consider Kapitza's PODF model given by Eq. (7.109).
a) Show that the mean implied by this PODF does not exist.
b) Show that the second moment about the center $t_c$ is infinite.

**7.4.3** A modification of the logistic model is given by the model of Schaefer

$$\frac{dP}{dt} = \frac{1}{\tau}\left(1 - \frac{P}{K}\right)P - E\,P.$$

The model, which was developed for the simulation of the development of fish populations, is equivalent to the logistic model for $E = 0$, where $L = P(-\infty) = 0$ is assumed for simplicity. The last term $-E\,P$ takes into account (human) predation that reduces the rate of population growth. It is reasonable to consider this term to be proportional to $P$: the effect of predation will increase with the population density. The variables $K$, $E < 1/\tau$, and $\tau$ are assumed to be non-negative and constant.
a) Write the model in the form of the logistic model (the structure of this rewritten model will be equal to the logistic model but the parameters are different).
b) Calculate the solution of this rewritten model by taking reference to the solution of the logistic model.
c) Explain the effect of a nonzero $E$ on the population dynamics in comparison to the logistic model.

**7.4.4** A modified version of the Schaefer model is given by the model

$$\frac{dP}{dt} = r\left(1 - \frac{P}{K}\right)P - H = -\frac{r}{K}\left(P - \frac{K}{2}\right)^2 + \frac{r\,K}{4} - H,$$

where $H$ is non-negative constant. The last expression provides a rewriting that is helpful for the following analysis.
a) Rewrite the model by using the nondimensional variables $P_* = P/K$ and $t_* = r\,t$.
b) Discuss the existence of equilibrium solutions for the cases $H > r\,K/4$, $H = r\,K/4$, and $H < r\,K/4$.
c) One of the three cases considered in b) leads to the conclusion that there are two equilibrium points. Write the nondimensional equation in the form $dP_*/dt_* = -(P_{*1} - P_*)(P_{*2} - P_*)$ for this case, where $P_{*1}$ and $P_{*2}$ refer to the two equilibrium states. Use this equation to discuss $dP_*/dt_*$ for the three cases $P_{*0} < P_{*1}$, $P_{*1} < P_{*0} < P_{*2}$, and $P_{*2} < P_{*0}$. Here, $P_{*0}$ denotes the initial value of $P_*$. Use this insight into the behavior of $dP_*/dt_*$ to explain the behavior of solutions close to the equilibrium points (explain which of the two equilibrium points $P_{*1}$ and $P_{*2}$ can be realized).

**7.4.5** The model of Gompertz, which is used for the modeling of the growth of a tumor, is given by the equation

$$\frac{dP}{dt} = \frac{1}{\tau}\ln\left(\frac{K}{P}\right)P.$$

Here, the variables $K$ and $\tau$ are assumed to be non-negative and constant.

a) Calculate the solution of this differential equation. Hint: use the variable transformation $y = \ln(P/K)$.

b) Calculate the equilibrium solution.

c) Show for $0 < P \le K$ that the change $dP/dt$ determined by the Gompertz model is equal or larger than $dP/dt = (1 - P/K)\,P/\tau$ as given by the logistic model. Hint: you may wish to show the validity and make use of $f(x) = 1 + \ln x - x \le 0$ for $x > 0$.

**7.4.6** The consideration of population development as a self-similar process does imply the following differential equation (Kapitza 1996),

$$\frac{dP}{P - P_c} = \alpha\frac{dt}{t - t_c}.$$

Here, $P_c$ and $t_c$ are the coordinates of a reference point, and $\alpha$ is a constant.

a) Find the solution to this differential equation. Express the constants in the solution in terms of the data $P_0$ and $t_0$ at the initial time.

b) The exponential growth and logistic models do not describe self-similar processes. Does this mean that the latter models have a poor standard?

**7.5.1** Consider a generalization of the differential equation (7.7),

$$\frac{dy}{dt} = -\frac{1}{T}\int_{-\infty}^{t}\mu(t-s)\left[y(s) - y_e\right]ds,$$

which includes memory effects. Here, $T$ and $y_e$ are constants, and $\mu(t-s)$ is a memory function. We assume that $\mu(t-s)$ is given by Eq. (7.120),

$$\mu(t-s) = \frac{1}{\tau_M}\exp\left\{-\frac{|\,t-s\,|}{\tau_M}\right\},$$

where $\tau_M$ refers to a characteristic memory time scale.

a) Write the differential equation for the limit case $\tau_M \to 0$.

b) Consider the case that $\tau_M$ is nonzero. Show that the differential equation considered can be written as (follow the discussion in Sect. 7.5.3)

$$\frac{d^2y}{dt^2} + \frac{1}{\tau_M}\frac{dy}{dt} + \frac{1}{\tau_M}\frac{y - y_e}{T} = 0.$$

c) Derive the limit $\tau_M \to 0$ of this second-order differential equation.

**7.5.2** Consider the second-order linear differential equation derived in part b) of the exercise 7.5.1.

a) Find the solution $y(t)$ to this equation. It is assumed that $\tau_M < T/4$.

b) Specify the eigenvalues for the case that $\tau_M << T$. Hint: you may wish to use the relation $(1 \pm x)^{1/2} = 1 \pm x/2$ for very small $x$.

c) Use the results obtained in b) to determine the solution for $\tau_M \to 0$.

**7.5.3** Consider again the differential equation discussed in exercise 7.5.1,

$$\frac{dy}{dt} = -\frac{1}{T}\int_{-\infty}^{t} \mu(t-s)\left[y(s) - y_e\right] ds\,.$$

For which choice of the memory function $\mu(t-s)$ does this equation imply the delay differential equation

$$\frac{dy}{dt} = -\frac{y(t-r) - y_e}{T}\,?$$

Here, $r$ refers to a non-negative and constant delay time.

# 8 Stochastic Evolution

Stochastic changes were discussed in Chap. 6 by involving randomness in the deterministic difference equations considered in Chap. 5. This approach can be used to demonstrate the relevance of randomness, to show typical formulations of stochastic equations, and to illustrate the analysis of stochastic processes. On the other hand, this approach is of empirical nature and does not reveal any insight into the general structure of equations for stochastic processes (i.e., the question of what are the laws of stochastics), the conditions for the applicability of certain equation types, the relations between equations for stochastic processes and equations for the PDF related to stochastic processes, and the solution of PDF equations. The latter questions will be addressed in this chapter. This will be done by considering the continuous evolution of stochastic processes, which represents an appropriate basis for explaining the relationship between equations for stochastic processes and differential equations for the PDF evolution. There is a close relation between this chapter and Chaps. 6 and 10. The difference equations described in Chap. 6 provide the numerical solution for the differential equations considered in this chapter, and the latter equations represent the basis for the stochastic differential equations for several variables considered in Chap. 10.

The motivation for considering the typical properties of stochastic differential equations and related PDF evolution equations will be explained in Sect. 8.1. Sections 8.2 and 8.3 will address the questions considered on the basis of PDF equations, i.e., evolution equations for PDFs and their solutions will be discussed. Sections 8.4 and 8.5 address the same questions on the basis of stochastic differential equations. Here, emphasis is placed on the relations between PDF equations and stochastic differential equations and the choice of an appropriate stochastic equation for the modeling of a stochastic process considered. The basic features of equations for continuous stochastic processes will be summarized in Sect. 8.6.

S. Heinz, *Mathematical Modeling*, DOI 10.1007/978-3-642-20311-4_8,

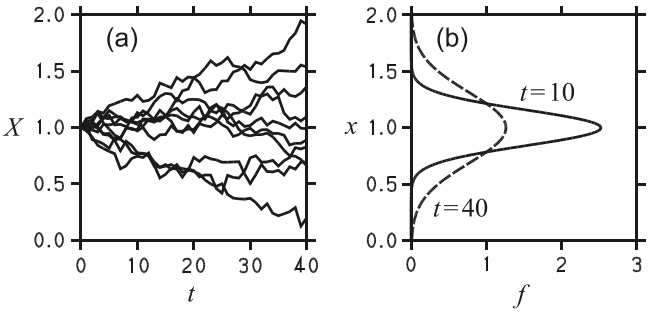



**Fig. 8.1.** The evolution of a stochastic process $X(t)$ in time $t$ is illustrated in terms of ten realizations in (a). The evolution of the corresponding PDF $f(x)$ is shown in (b) at $t = 10$ (*solid line*) and $t = 40$ (*dashed line*).

## 8.1 Motivation

**Stochastic Process and PDF Evolution.** There are two possibilities to look at the evolution of a random variable and its PDF. The first approach, which was applied in Chap. 6, is to consider an evolution equation for the random variable. This approach determines the evolution of the corresponding PDF: the PDF can be calculated numerically at every time (see the illustration in Fig. 8.1). The sec-ond approach is to consider an evolution equation for the PDF of a random variable considered. Such a PDF equation determines the evolution of a stochastic process: according to Sect. 4.2.3 we can represent the PDF at every time in terms of random numbers (the correlations of a stochastic process are also determined by the PDF equation). The availability of a PDF evolution equation appears to be very helpful because of several reasons:

- The PDF evolution equation determines an evolution of the stochastic process considered. Therefore, we can use the PDF evolution equation to validate and possibly to generalize the concepts used in Chap. 6 for the description of the evolution of a stochastic process.
- The numerical PDF calculation is always affected by the number of realizations applied and the numerical method used for calculating the filtered PDF. On the other hand, the calculation of a PDF via its evolution equation can provide a PDF that is unaffected by the number of samples (and the filter width) applied.
- The existence of a PDF evolution equation offers the chance to find analytically the PDF, which would imply a significant reduction of the problem considered (such that there is possibly only the need to determine the mean and variance of random variables).

- The existence of a PDF evolution equation offers the chance to find evolution equations for the mean, variance, and other moments of the process considered. Such moment equations can be solvable such that the temporal development of moments can be determined.

**Questions Considered.** To take advantage of a PDF equation, we will derive in this chapter a PDF equation and analyze the relationship between equations for a stochastic process and for its PDF. In particular, we will consider the following questions:

- Which equation describes the evolution of the PDF of a stochastic process?
- How can we solve this PDF equation?
- Which stochastic process equation corresponds to this PDF equation?
- How can we determine stochastic equations for the modeling of any case?

The derivation and solution of equations for moments (means and variances) of a PDF will be addressed, too. The latter questions will be discussed on the basis of differential equations for the evolution of both the PDF and the stochastic process considered, which simplifies the explanations significantly. In this chapter we focus on equations for one random variable, such that the characteristic features can be explained in a relatively simple way. The extension of the single-variable concepts to the case of several variables will be presented in Chap. 10.

## 8.2 PDF Evolution Equations

Let us address first the question of how the evolution of a PDF in time can be described. First, we will determine the general structure of an equation for the PDF of any stochastic process. Second, we will ask under which conditions this general equation can be simplified to an equation that is useful for the modeling of stochastic processes.

### *8.2.1 The Kramers-Moyal Equation*

**PDF Definition.** The PDF of a random variable $X$ is defined by the expression $f(x) = <\delta(x - X)>$. Here, $\delta(x - X)$ refers to a delta function (see the explanations in Sect. 4.2.2). The expression $f(x) = <\delta(x - X)>$ also can be used for a stochastic process that changes in time. The PDF $f(x, t)$ at the time $t$ is then defined by

$$f(x,t) = \left\langle \delta\left(x - X(t)\right)\right\rangle. \tag{8.1}$$

At the later time $t + \Delta t$, the PDF is given by

$$f(x,t+\Delta t) = \left\langle \delta\big(x - X(t+\Delta t)\big)\right\rangle. \tag{8.2}$$

There is no need to make any assumption about the time interval $\Delta t$ here: it can be short or not. In order to calculate the evolution of $f(x, t)$ in time, we have to relate the PDF $f(x, t + \Delta t)$ at the later time to the PDF $f(x, t)$ at the previous time. This is the question that will be considered in the following.

**Kramers-Moyal Equation.** The best way to address the latter question is to consider first the instantaneous PDF involved in Eq. (8.2),

$$\delta\big(x - X(t+\Delta t)\big) = \delta(z). \tag{8.3}$$

For simplicity, the abbreviation $z = x - X(t + \Delta t)$ is introduced here. To relate the instantaneous PDF at $t + \Delta t$ to the instantaneous PDF at $t$, we consider the Taylor expansion of $\delta(z)$ at $z_0 = x - X(t)$. This expansion reads

$$\delta(z) = \sum_{n=0}^{\infty} \frac{\delta^{(n)}(z_0)}{n!}(z - z_0)^n, \tag{8.4}$$

where $\delta^{(n)}$ refers to the $n^{\text{th}}$-order derivative of $\delta(z_0)$. The latter expression also can be written

$$\delta(z) = \delta(z_0) + \sum_{n=1}^{\infty}\left(\frac{d}{dx}\right)^n\left[\frac{(-1)^n}{n!}(z_0 - z)^n\,\delta(z_0)\right]. \tag{8.5}$$

This expression results from the explicit consideration of the contribution at $n = 0$, the $n^{\text{th}}$-order derivative of $\delta(z_0)$ is written in terms of the corresponding derivative by $x$, and the rewriting $(z - z_0)^n = (-1)^n (z_0 - z)^n$ is used. The derivatives apply to all the bracket terms, but the only term that depends on $x$ is $\delta(z_0)$ because $z_0 - z = X(t + \Delta t) - X(t)$ is independent of $x$. By averaging Eq. (8.5) we obtain

$$f(x,t+\Delta t) = f(x,t) + \sum_{n=1}^{\infty}\left(-\frac{d}{dx}\right)^n \frac{\left\langle (z_0 - z)^n\,\delta(z_0)\right\rangle}{n!}, \tag{8.6}$$

where the definitions of $f(x, t + \Delta t)$ and $f(x, t)$ are used. We write the term $f(x, t)$ on the left-hand side, divide both sides by $\Delta t$, and take the limit $\Delta t \to 0$,

$$\lim_{\Delta t\to 0}\frac{f(x,t+\Delta t) - f(x,t)}{\Delta t} = \sum_{n=1}^{\infty}\left(-\frac{d}{dx}\right)^n \lim_{\Delta t\to 0}\frac{\left\langle (z_0 - z)^n\,\delta(z_0)\right\rangle}{n!\,\Delta t}. \tag{8.7}$$

The left-hand side represents the partial derivative of $f$ by $t$. Hence, we have

$$\frac{\partial f(x,t)}{\partial t} = \sum_{n=1}^{\infty}\left(-\frac{\partial}{\partial x}\right)^n D^{(n)}(x,t)\,f(x,t), \tag{8.8}$$

where we used the abbreviation

$$D^{(n)}(x,t) = \lim_{\Delta t \to 0} \frac{\left\langle (z_0 - z)^n \delta(z_0) \right\rangle}{n! \Delta t \, f(x,t)}. \tag{8.9}$$

The derivatives by $x$ are written now as partial derivatives. The latter derivatives apply to the product $D^{(n)}(x, t) f(x, t)$. Equation (8.8) is referred to as the Kramers-Moyal equation (Kramers 1940, Moyal 1949), and $D^{(n)}(x, t)$ denote the Kramers-Moyal coefficients.

**Kramers-Moyal Coefficients.** The Kramers-Moyal coefficients $D^{(n)}(x, t)$ can be rewritten by using the definitions $z_0 - z = X(t + \Delta t) - X(t)$ and $z_0 = x - X(t)$,

$$D^{(n)}(x,t) = \lim_{\Delta t \to 0} \frac{\left\langle \big(X(t+\Delta t) - X(t)\big)^n \delta(x - X(t)) \right\rangle}{n! \Delta t \, f(x,t)}. \tag{8.10}$$

The properties of delta functions imply that $\delta(x - X(t))$ is zero except at $X(t) = x$. A convenient way to account for the condition $X(t) = x$ is to use the definition of a conditional mean (see the corresponding explanations in Sect. 10.2.2) as an abbreviation. Such a conditional mean is defined for any function $g(X(t))$ by

$$\frac{1}{f(x,t)} \left\langle g\big(X(t)\big) \, \delta(x - X(t)) \right\rangle = \left\langle g\big(X(t)\big) \,|\, X(t) = x \right\rangle = \left\langle g\big(X(t)\big) \,|\, x, t \right\rangle. \tag{8.11}$$

The last expression represents a convenient rewriting of the previous expression: it refers to the condition that $X = x$ at $t$. In terms of this definition, the Kramers-Moyal coefficients can be written

$$D^{(n)}(x,t) = \lim_{\Delta t \to 0} \frac{\left\langle \big(X(t+\Delta t) - X(t)\big)^n \,|\, x, t \right\rangle}{n! \Delta t}. \tag{8.12}$$

**Markov Process.** The difference $\Delta X = X(t + \Delta t) - X(t)$ in the definition of $D^{(n)}$ may depend on all the values of the random variable $X$ at earlier times, this means on $X(t - k\,\Delta t)$ with $k = 0, 1, \ldots$. Depending on an appropriate choice of the random variable considered, it is relatively often the case that the influence of such memory effects is relatively small, this means that $\Delta X$ can be considered to be determined by the state $X(t)$. Stochastic processes for which $\Delta X$ does only depend on the previous state $X(t)$ are referred to as Markov processes (Gardiner 1983, Risken 1984). Such Markov processes will be considered from now, which simplifies the application of the Kramers-Moyal equation significantly. For this case, the coefficients $D^{(n)}(x, t)$ are only functions of $x$ and $t$. The Kramers-Moyal equation (8.8) then represents a differential equation of first-order with respect to time $t$. Combined with appropriate boundary conditions and the specification of an initial PDF $f(x, t_0)$, Eq. (8.8) uniquely determines the PDF $f(x, t)$.

### *8.2.2 The Pawula Theorem*

**Pawula's Theorem.** The Kramers-Moyal equation (8.8) involves an infinite number of coefficients. Regarding the use of this equation it is, therefore, relevant to know how many coefficients should be taken into account. A relevant requirement for dealing with the latter question arises from Pawula's theorem, which will be presented next. This theorem can be obtained like Schwarz's inequality (see the discussion of the correlation coefficient in Sect. 2.3.1). We introduce a non-negative function $H(p) \geq 0$ as

$$\begin{aligned} H(p) &= \left\langle \left(\Delta X^k + p\,\Delta X^{k+m}\right)^2 \mid x,t \right\rangle \\ &= \left\langle \Delta X^{2k} \mid x,t \right\rangle + 2p\left\langle \Delta X^{2k+m} \mid x,t \right\rangle + p^2 \left\langle \Delta X^{2k+2m} \mid x,t \right\rangle. \end{aligned} \tag{8.13}$$

Here, $\Delta X = X(t + \Delta t) - X(t)$, and we assume that $k \geq 1$ and $m \geq 0$. The first two derivatives of $H(p)$ by $p$ are given by

$$\frac{dH}{dp} = 2\left\langle \Delta X^{2k+m} \mid x,t \right\rangle + 2p\left\langle \Delta X^{2k+2m} \mid x,t \right\rangle, \tag{8.14a}$$

$$\frac{d^2H}{dp^2} = 2\left\langle \Delta X^{2k+2m} \mid x,t \right\rangle. \tag{8.14b}$$

These two derivatives show that $H(p)$ has a minimum at

$$p_c = -\frac{\left\langle \Delta X^{2k+m} \mid x,t \right\rangle}{\left\langle \Delta X^{2k+2m} \mid x,t \right\rangle}. \tag{8.15}$$

The minimum $H_{\min}$ of $H(p)$ is given by

$$H_{\min} = \left\langle \Delta X^{2k} \mid x,t \right\rangle - \frac{\left\langle \Delta X^{2k+m} \mid x,t \right\rangle^2}{\left\langle \Delta X^{2k+2m} \mid x,t \right\rangle}. \tag{8.16}$$

The function $H(p) \geq 0$ for all $p$, this means we do also have $H_{\min} \geq 0$. Therefore, Eq. (8.16) implies that

$$\left\langle \Delta X^{2k} \mid x,t \right\rangle \left\langle \Delta X^{2k+2m} \mid x,t \right\rangle \geq \left\langle \Delta X^{2k+m} \mid x,t \right\rangle^2. \tag{8.17}$$

This condition is not useful for $m = 0$. Thus, we can consider $m \geq 1$ from now. By dividing Eq. (8.17) by $(\Delta t)^2$, taking the limit $\Delta t \to 0$, and using Eq. (8.12) of the Kramers-Moyal coefficients $D^{(n)}(x, t)$, we find that

$$(2k)!D^{(2k)}\,(2k+2m)!D^{(2k+2m)} \geq \left[(2k+m)!D^{(2k+m)}\right]^2. \tag{8.18}$$

For simplicity, the dependence of $D^{(n)}$ on $x$ and $t$ is not indicated. Equations (8.18) represent the Theorem of Pawula (1967).

**Consequences of Pawula's Theorem.** The consequences of Pawula's theorem can be seen by considering the cases that $D^{(2k)} = 0$ and $D^{(2k+2m)} = 0$, respectively. The right-hand side of Eq. (8.18) is positive, and it has to be smaller or equal than zero for $D^{(2k)} = 0$ and $D^{(2k+2m)} = 0$. The latter constraints imply that

$$D^{(2k)} = 0 \quad \rightarrow \quad D^{(2k+m)} = 0, \tag{8.19a}$$

$$D^{(2k+2m)} = 0 \quad \rightarrow \quad D^{(2k+m)} = D^{(2k+2m-m)} = 0. \tag{8.19b}$$

These two conditions can be simplified to one condition for any $n \geq 1$,

$$D^{(2n)} = 0 \quad \rightarrow \quad D^{(3)} = D^{(4)} = D^{(5)} = \cdots D^{(\infty)} = 0. \tag{8.20}$$

The coefficients $D^{(1)}$ and $D^{(2)}$ are unaffected by this constraint because $2k + m$ cannot be smaller than 3 for $k \geq 1$ and $m \geq 1$. Thus, we do only have two options depending on whether or not the consequence of Eq. (8.20) does apply.

a) All even coefficients $D^{(2n)}$ for $n \geq 1$ are nonzero. The consequence of Eq. (8.20) does not apply then. We have to deal with an infinite number of coefficients.
b) Any even coefficient $D^{(2n)}$ is zero, where $n \geq 1$. The consequence of Eq. (8.20) does apply then. This means that we have $D^{(3)} = D^{(4)} = D^{(5)} = \cdots = D^{(n)} = 0$.

What is the consequence if a Kramers-Moyal equation is used that is not consistent with Pawula's theorem (e.g., an equation that involves nonzero coefficients up to fourth order)? Pawula's theorem corresponds to the consideration of the quadratic moment $H(p)$ as a non-negative number. Thus, we may find negative squared moments if Pawula's theorem is not satisfied. Moments represent integrals over the corresponding PDF. Hence, the PDF must have negative values if Pawula's theorem is disregarded.

### *8.2.3 The Fokker-Planck Equation*

**Fokker-Planck Equation.** The consideration of an infinite number of even Kramers-Moyal coefficients implies the problem of providing all the coefficients as functions of $x$. Apart from that, an equation with an infinite number of terms is difficult to treat numerically. Hence, we will apply the option b) described in the preceding paragraph, i.e., we neglect coefficients $D^{(n)}$ with $n \geq 3$. Hence, we will consider from now on the equation

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial D^{(1)}(x,t) f(x,t)}{\partial x} + \frac{\partial^2 D^{(2)}(x,t) f(x,t)}{\partial x^2}. \tag{8.21}$$

This equation represents the Fokker-Planck equation (Fokker 1914, Planck 1917).

According to Eq. (8.12), $D^{(1)}$ and $D^{(2)}$ are defined by

$$D^{(1)}(x,t) = \lim_{\Delta t \to 0} \frac{\langle \Delta X \mid x,t \rangle}{\Delta t}, \qquad D^{(2)}(x,t) = \lim_{\Delta t \to 0} \frac{\langle \Delta X^2 \mid x,t \rangle}{2\Delta t}. \tag{8.22}$$

Under which condition is the neglect of $D^{(n)}$ with $n \geq 3$ justified? The answer to this question is closely related to the consideration of the continuity of the sample path of stochastic processes. By considering an infinitesimal time increment $\Delta t$ it is relatively often the case that the change $\Delta X = X(t + \Delta t) - X(t)$ is bounded (i.e., relatively small). Such stochastic processes have a continuous sample path. The consideration of such a process implies the neglect of $D^{(n)}$ with $n \geq 3$ (see Gardiner 1983). In other words, jump processes (instantaneous changes $\Delta X$ that may be very large) with a discontinuous sample path are taken into account if coefficients $D^{(n)}$ with $n \geq 3$ are involved.

**Mean Equation.** The implications of the Fokker-Planck equation (8.21) for the evolution of moments of the PDF $f(x, t)$ will be considered next. A main purpose of this discussion is to provide insight into the relevance of the model parameters of the Fokker-Planck equation. By multiplying the Fokker-Planck equation (8.21) by $x$ and integration over the sample space we obtain

$$\int x \frac{\partial f(x,t)}{\partial t} dx = -\int x \frac{\partial D^{(1)}(x,t) f(x,t)}{\partial x} dx + \int x \frac{\partial^2 D^{(2)}(x,t) f(x,t)}{\partial x^2} dx. \tag{8.23}$$

This equation has to be rewritten in terms of moments of $f(x, t)$, i.e., as functions of $x$ multiplied by the PDF $f(x, t)$ and integrated over the sample space. To find the corresponding expressions we write

$$\begin{aligned} \frac{\partial}{\partial t} \int x f(x,t)\, dx &= -\int \frac{\partial x D^{(1)}(x,t) f(x,t)}{\partial x} dx + \int \frac{\partial x}{\partial x} D^{(1)}(x,t) f(x,t)\, dx \\ &\quad + \int \frac{\partial}{\partial x} \left[ x \frac{\partial D^{(2)}(x,t) f(x,t)}{\partial x} \right] dx - \int \frac{\partial x}{\partial x} \frac{\partial D^{(2)}(x,t) f(x,t)}{\partial x} dx \\ &= -\left( x D^{(1)}(x,t) f(x,t) \right)_{-\infty}^{\infty} + \int D^{(1)}(x,t) f(x,t)\, dx \\ &\quad + \left( x \frac{\partial D^{(2)}(x,t)}{\partial x} f(x,t) + x D^{(2)}(x,t) \frac{\partial f(x,t)}{\partial x} \right)_{-\infty}^{\infty} - \left( D^{(2)}(x,t) f(x,t) \right)_{-\infty}^{\infty}. \end{aligned} \tag{8.24}$$

The integral on the left-hand side is equal to $\langle X \rangle$, and the second integral on the right-hand side represents $\langle D^{(1)} \rangle$. The other integrals on the right-hand side do vanish because the PDF $f(x, t)$ and its derivatives are zero at infinity. Therefore, Eq. (8.24) can be written

$$\frac{d\langle X \rangle}{dt} = \langle D^{(1)} \rangle. \tag{8.25}$$

The partial derivative by $t$ can be replaced here by the normal derivative because $\langle X\rangle$ and $\langle D^{(1)}\rangle$ are only functions of $t$. Hence, $\langle D^{(1)}\rangle$ determines the transport of the means $\langle X\rangle$. For that reason, $D^{(1)}$ is called the drift coefficient.

**Variance Equation.** The variance equation can be obtained similarly. We multiply the Fokker-Planck equation (8.21) by $x^2$ and integrate over the sample space,

$$\int x^2 \frac{\partial f(x,t)}{\partial t} dx = -\int x^2 \frac{\partial D^{(1)}(x,t) f(x,t)}{\partial x} dx + \int x^2 \frac{\partial^2 D^{(2)}(x,t) f(x,t)}{\partial x^2} dx. \tag{8.26}$$

In correspondence to the rewriting of Eq. (8.23) we write

$$\begin{aligned} \frac{\partial}{\partial t}\int x^2 f(x,t)\, dx = &-\int \frac{\partial x^2\, D^{(1)}(x,t) f(x,t)}{\partial x}\, dx + \int \frac{\partial x^2}{\partial x} D^{(1)}(x,t) f(x,t)\, dx \\ &+ \int \frac{\partial}{\partial x}\left[ x^2 \frac{\partial D^{(2)}(x,t) f(x,t)}{\partial x} \right] dx - \int \frac{\partial x^2}{\partial x} \frac{\partial D^{(2)}(x,t) f(x,t)}{\partial x}\, dx. \end{aligned} \tag{8.27}$$

The integral on the left-hand side is equal to $\langle X^2\rangle$. The first and third term on the right-hand side disappear as the corresponding terms in Eq. (8.24). Thus,

$$\begin{aligned} \frac{\partial \langle X^2 \rangle}{\partial t} &= 2\int x D^{(1)}(x,t) f(x,t)\, dx - 2\int x \frac{\partial D^{(2)}(x,t) f(x,t)}{\partial x}\, dx \\ &= 2\langle X D^{(1)} \rangle - 2\int \frac{\partial x D^{(2)}(x,t) f(x,t)}{\partial x}\, dx + 2\int \frac{\partial x}{\partial x} D^{(2)}(x,t) f(x,t)\, dx. \end{aligned} \tag{8.28}$$

The second term of the last line is equal to zero because we take an integral over a derivative. By rewriting the last integral we obtain

$$\frac{d\langle X^2 \rangle}{dt} = 2\langle X D^{(1)} \rangle + 2\langle D^{(2)} \rangle. \tag{8.29}$$

Here, the partial derivative by $t$ is replaced by the normal derivative by $t$. Instead of considering the equations for second-order moments, it is more convenient to consider equations for the variance

$$\langle \widetilde{X}^2 \rangle = \langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2. \tag{8.30}$$

By differentiating this variance expression by $t$ we obtain

$$\frac{d\langle \widetilde{X}^2 \rangle}{dt} = \frac{d\langle X^2 \rangle}{dt} - \frac{d\langle X \rangle^2}{dt} = \frac{d\langle X^2 \rangle}{dt} - 2\langle X \rangle \frac{d\langle X \rangle}{dt}. \tag{8.31}$$

The use of Eqs. (8.29) and (8.25) then implies the following variance equation,

$$\frac{d\langle \widetilde{X}^2 \rangle}{dt} = 2\langle X D^{(1)} \rangle + 2\langle D^{(2)} \rangle - 2\langle X \rangle \langle D^{(1)} \rangle = 2\langle \widetilde{X}\, \widetilde{D}^{(1)} \rangle + 2\langle D^{(2)} \rangle. \tag{8.32}$$

In the last expression we used $< \widetilde{X}\, \widetilde{D}^{(1)} > = <X\, D^{(1)}> - <X>\, <D^{(1)}>$, which may be proven by distributing the left-hand side. Hence, variance is produced by $<D^{(2)}>$ if $D^{(2)}$ is positive. The corresponding increase of the width of the PDF characterizes a diffusion process, which is the reason why $D^{(2)}$ is called a diffusion coefficient. An equilibrium state may be reached asymptotically if the first term on the right-hand side of Eq. (8.32) appears with a negative sign, i.e., if this terms describes a dissipation (or destruction) of variance.

**Correlations.** The correlation between $X(t)$ and $X(t')$ can be calculated on the basis of the Fokker-Planck equation (8.21). Without loss of generality we assume that $t \le t' = t + r$, where $r$ is any non-negative time. As shown in exercise 8.3.1, the correlation between $X(t)$ and $X(t + r)$ is found to be determined by the equation

$$\frac{d\left\langle \widetilde{X}(t)\widetilde{X}(t+r)\right\rangle}{dr} = \left\langle \widetilde{X}(t)\,\widetilde{D}^{(1)}\big(X(t+r),t+r\big)\right\rangle. \tag{8.33}$$

The correlation at $r = 0$ is determined by the variance equation (8.32). We see that the correlation is unaffected by the diffusion coefficient $D^{(2)}$. Thus, correlations are not produced, but correlations relax according to the model provided by $D^{(1)}$.

## 8.3 Solution to the Fokker-Planck Equation

Let us illustrate basic characteristics of solutions to the Fokker-Planck equation (8.21) by considering the following example, which enables the calculation of an analytical solution. The equation considered is given by

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}\Big[F(t) + G(t)\big(x - \langle X\rangle\big)\Big] f(x,t) + \frac{\partial^2 D(t)\, f(x,t)}{\partial x^2}. \tag{8.34}$$

The drift coefficient $D^{(1)}$ is a linear function of $x$. The inclusion of $<X>$ in $D^{(1)}$ represents a convenient writing: the term $G <X>$ could be also combined with the drift term $F$. Such a linear model for $D^{(1)}$ is well appropriate to characterize near-equilibrium processes. The diffusion coefficient $D(t)$ is assumed to depend only on $t$, which is often a convenient assumption. Equation (8.34) will be combined with natural boundary conditions, this means we have $f(x, t) \to 0$ for $|x| \to \infty$.

### *8.3.1 The Solution Approach*

**Solution Approach.** The solutions $f(x, t)$ to the Fokker-Planck equation (8.34) involve (*i*) information about the initial PDF $f(x_0, t_0)$, and (*ii*) information about the

transition from the initial PDF to any asymptotic PDF, which is determined by the PDF evolution equation. It is very helpful to represent the solution such that these two ingredients are separated from each other: this simplifies significantly the use of the solution for any initial PDFs. The latter can be achieved in the following way. First, we represent the PDF $f(x)$ as

$$f(x,t) = \int f(x,t;x_0,t_0)\, dx_0 . \tag{8.35}$$

Here, $f(x, t, x_0, t_0)$ represents the two-point PDF, which is defined by

$$f(x,t;x_0,t_0) = \left\langle \delta(x - X(t))\, \delta(x_0 - X(t_0)) \right\rangle . \tag{8.36}$$

The consistency of Eq. (8.35) can be seen by using the definition (8.36) in Eq. (8.35) and performing the integration. This recovers the definition (8.1) of $f(x, t)$. Second, we introduce the conditional PDF $f(x, t \mid x_0, t_0)$ by the relation

$$\begin{aligned} f(x,t\,|\,x_0,t_0) &= \frac{f(x,t;x_0,t_0)}{f(x_0,t_0)} = \frac{\left\langle \delta(x - X(t))\, \delta(x_0 - X(t_0)) \right\rangle}{\left\langle \delta(x_0 - X(t_0)) \right\rangle} \\ &= \left\langle \delta(x - X(t)) \,|\, X(t_0) = x_0 \right\rangle = \left\langle \delta(x - X(t)) \,|\, x_0, t_0 \right\rangle . \end{aligned} \tag{8.37}$$

The last expression of the first line results from the use of the definitions of the two-point PDF $f(x, t, x_0, t_0)$ and initial PDF $f(x_0, t_0)$. The expressions in the last line make use of the definition (8.11) of conditional means. Hence, $f(x, t \mid x_0, t_0)$ is the PDF of $X(t)$ under the condition that $X(t_0) = x_0$ (see also the discussions related to conditional PDFs in Sect. 10.2.2). By relating the two-point PDF to the conditional PDF according to Eq. (8.37) we can write Eq. (8.35) as

$$f(x,t) = \int f(x,t\,|\,x_0,t_0)\; f(x_0,t_0)\, dx_0 . \tag{8.38}$$

In this way, the PDF $f(x, t)$ is calculated as an integral over the conditional PDF $f(x, t \mid x_0, t_0)$ multiplied by the probability for having the initial value $x_0$. The advantage of this approach is that we can calculate a general expression for the conditional PDF $f(x, t \mid x_0, t_0)$, which is independent of the initial PDF $f(x_0, t_0)$. The PDF $f(x, t)$ can be obtained then for any initial PDF $f(x_0, t_0)$ by integration according to Eq. (8.38).

**Conditional PDF Equation.** How is it possible to find the conditional PDF $f(x, t \mid x_0, t_0)$? Using Eq. (8.38), the Fokker-Planck equation (8.34) can be written

$$\begin{aligned} \int \left\{ \frac{\partial f(x,t\,|\,x_0,t_0)}{\partial t} + \frac{\partial \left[ F(t) + G(t)\left(x - \langle X \rangle\right) \right] f(x,t\,|\,x_0,t_0)}{\partial x} - \frac{\partial^2 D(t) f(x,t\,|\,x_0,t_0)}{\partial x^2} \right\} \\ \cdot f(x_0,t_0)\, dx_0 = 0 . \end{aligned} \tag{8.39}$$

Hence, the conditional PDF $f(x, t \mid x_0, t_0)$ has to satisfy the Fokker-Planck equation (8.34), too, this means the conditional PDF $f(x, t \mid x_0, t_0)$ has to satisfy the equation

$$\frac{\partial f(x,t\,|\,x_0,t_0)}{\partial t} = -\frac{\partial\left[F(t)+G(t)\left(x-\langle X\rangle\right)\right]f(x,t\,|\,x_0,t_0)}{\partial x} + \frac{\partial^2 D(t) f(x,t\,|\,x_0,t_0)}{\partial x^2}. \tag{8.40}$$

**Conditional PDF Initial Condition.** The initial condition that is required for the calculation of the conditional PDF $f(x, t \mid x_0, t_0)$ can be derived from Eq. (8.37),

$$\begin{aligned} f(x,t_0\,|\,x_0,t_0) &= \frac{\langle \delta(x-X(t_0))\,\delta(x_0-X(t_0))\rangle}{\langle \delta(x_0-X(t_0))\rangle} = \frac{\langle \delta(x-x_0)\,\delta(x_0-X(t_0))\rangle}{\langle \delta(x_0-X(t_0))\rangle} \\ &= \delta(x-x_0). \end{aligned} \tag{8.41}$$

Here, we used the definition of $f(x, t \mid x_0, t_0)$ at $t = t_0$ and the sifting property of the delta function to obtain the last expression.

## *8.3.2 The Solution to the Fokker-Planck Equation*

**Conditional PDF Model.** The asymptotic conditional PDF $f(x, t \mid x_0, t_0)$ that is implied by the Fokker-Planck equation (8.40) is a normal PDF: see exercise 8.3.2. In addition, a normal PDF can satisfy the delta function initial condition (8.41) for the conditional PDF (see exercise 8.3.3). Therefore, it makes sense to ask whether the conditional PDF can be described in each instance by a normal PDF. To prove the suitability of this idea, we model the conditional PDF $f(x, t \mid x_0, t_0)$ by a normal PDF,

$$f(x,t\,|\,x_0,t_0) = \frac{1}{\sqrt{2\pi\beta}}\exp\left\{-\frac{(x-\alpha)^2}{2\beta}\right\}. \tag{8.42}$$

Here, $\alpha(t)$ represents the mean value and $\beta(t)$ is the variance of this conditional PDF model. To be consistent with the initial condition (8.41), $\alpha$ and $\beta$ need to have the initial values $\alpha(t_0) = x_0$ and $\beta(t_0) = 0$ (see exercise 8.3.3). Another view of looking at the assumption (8.42) is the following: we ask under which conditions we may have a normal PDF as solution of the Fokker-Planck equation. By using expression (8.42), the solution of the Fokker-Planck equation (8.34) is given by

$$f(x,t) = \int \frac{1}{\sqrt{2\pi\beta}}\exp\left\{-\frac{(x-\alpha)^2}{2\beta}\right\} f(x_0,t_0)\,dx_0. \tag{8.43}$$

Here, $f(x_0, t_0)$ can be any initial PDF. It is worth noting that $\alpha$ depends on $x_0$ via its initial value $\alpha(t_0) = x_0$.

**Conditional PDF Model Validity.** To prove the suitability of the assumption (8.42) we have to show that Eq. (8.42) can satisfy the Fokker-Planck equation (8.40). To address this question we write Eq. (8.40) as

$$\frac{\partial f(x,t\,|\,x_0,t_0)/\partial t}{f(x,t\,|\,x_0,t_0)} = -G(t) - \left[F(t) + G(t)\left(x - \langle X\rangle\right)\right]\frac{\partial f(x,t\,|\,x_0,t_0)/\partial x}{f(x,t\,|\,x_0,t_0)} + D(t)\,\frac{\partial^2 f(x,t\,|\,x_0,t_0)/\partial x^2}{f(x,t\,|\,x_0,t_0)}. \tag{8.44}$$

The derivatives required in this relation follow from Eq. (8.42),

$$\frac{\partial f(x,t\,|\,x_0,t_0)/\partial t}{f(x,t\,|\,x_0,t_0)} = -\frac{\beta'}{2\beta} + \frac{(x-\alpha)^2}{2\beta^2}\beta' + \frac{x-\alpha}{\beta}\alpha', \tag{8.45a}$$

$$\frac{\partial f(x,t\,|\,x_0,t_0)/\partial x}{f(x,t\,|\,x_0,t_0)} = -\frac{x-\alpha}{\beta}, \tag{8.45b}$$

$$\frac{\partial^2 f(x,t\,|\,x_0,t_0)/\partial x^2}{f(x,t\,|\,x_0,t_0)} = \frac{(x-\alpha)^2}{\beta^2} - \frac{1}{\beta}. \tag{8.45c}$$

Here, $\alpha'$ and $\beta'$ refer to the derivatives of $\alpha$ and $\beta$ by $t$, respectively. By making use of these derivatives in Eq. (8.44) we find the relation

$$-\frac{\beta'}{2\beta} + \frac{(x-\alpha)^2}{2\beta^2}\beta' + \frac{x-\alpha}{\beta}\alpha' = \\ = -G(t) + \left[F(t) + G(t)\left(x - \alpha + \alpha - \langle X\rangle\right)\right]\frac{x-\alpha}{\beta} + D(t)\left[\frac{(x-\alpha)^2}{\beta^2} - \frac{1}{\beta}\right]. \tag{8.46}$$

Here, $-\alpha + \alpha$ was added to the second term on the right-hand side to simplify the following derivations. By considering terms that are independent of $x - \alpha$, linear in $x - \alpha$, and quadratic in $x - \alpha$, the latter condition can be written

$$0 = \left(\frac{(x-\alpha)^2}{2\beta^2} - \frac{1}{2\beta}\right)\left[\beta' - 2\beta G(t) - 2D(t)\right] + \frac{x-\alpha}{\beta}\left[\alpha' - G(t)\left(\alpha - \langle X\rangle\right) - F(t)\right]. \tag{8.47}$$

The terms that are independent of $x - \alpha$ and quadratic are combined here because they have the same bracket term. This condition has to be satisfied for all $x$. The resulting requirement is that the two bracket terms have to disappear. We obtain two equations for $\alpha$ and $\beta$ in this way,

$$\frac{d\alpha}{dt} = G(t)\left(\alpha - \langle X\rangle\right) + F(t), \tag{8.48a}$$

$$\frac{d\beta}{dt} = 2G(t)\beta + 2D(t). \tag{8.48b}$$

The fact that it is possible to satisfy the condition (8.47) for all $x$ means that the model (8.42) represents an appropriate model for the conditional PDF, i.e., this model satisfies the Fokker-Planck equation (8.40) for the conditional PDF.

### *8.3.3 Process and Conditional PDF Statistics*

**Relations Between Process and Conditional PDF Statistics.** It is interesting to compare the Eqs. (8.48) for the mean $\alpha$ and variance $\beta$ of the conditional PDF model (8.42) with the equations for the mean $<X>$ and the variance $< \widetilde{X}^2 >$, which follow from the Fokker-Planck equation (8.34). The simplest way to obtain the latter equations is to use the general Eqs. (8.25) and (8.32) for the mean $<X>$ and variance $< \widetilde{X}^2 >$ for the case considered,

$$\frac{d\left\langle X\right\rangle}{dt} = \left\langle D^{(1)}\right\rangle = \left\langle F(t) + G(t)\ \widetilde{X}\right\rangle = F(t). \tag{8.49a}$$

$$\begin{aligned}\frac{d\left\langle \widetilde{X}^2\right\rangle}{dt} &= 2\left\langle \widetilde{X}\ \widetilde{D}^{(1)}\right\rangle + 2\left\langle D^{(2)}\right\rangle = 2\left\langle \widetilde{X}\left(F(t) + G(t)\ \widetilde{X}\right)\right\rangle + 2\,D(t) \\ &= 2G(t)\left\langle \widetilde{X}^2\right\rangle + 2\,D(t).\end{aligned} \tag{8.49b}$$

Equations (8.49) reveal a significant similarity to Eqs. (8.48) for the model parameters $\alpha$ and $\beta$ of the conditional PDF. To see the difference in more detail, let us use Eqs. (8.48) and (8.49) to find the differences $\alpha - <X>$ and $\beta - < \widetilde{X}^2 >$,

$$\frac{d}{dt}\left(\alpha - \left\langle X\right\rangle\right) = G(t)\left(\alpha - \left\langle X\right\rangle\right), \tag{8.50a}$$

$$\frac{d}{dt}\left(\beta - \left\langle \widetilde{X}^2\right\rangle\right) = 2\,G(t)\left(\beta - \left\langle \widetilde{X}^2\right\rangle\right). \tag{8.50b}$$

The solutions of these equations, which satisfy the initial conditions $\alpha(t_0) = x_0$ and $\beta(t_0) = 0$, are given by

$$\alpha = \left\langle X\right\rangle + \left(x_0 - \left\langle X_0\right\rangle\right)\exp\left\{\int_{t_0}^{t} G(s)\,ds\right\}, \tag{8.51a}$$

$$\beta = \left\langle \widetilde{X}^2\right\rangle - \left\langle \widetilde{X}_0{}^2\right\rangle \exp\left\{2\int_{t_0}^{t} G(s)\,ds\right\}. \tag{8.51b}$$

For a positive $G$, these solutions imply model parameters $\alpha$ and $\beta$ that approach infinity asymptotically – which does not make sense. Consequently, $G$ has to be negative.

**Asymptotic Evolution.** For a negative $G(t)$, $\alpha$ and $\beta$ relax asymptotically (i.e., for $t \to \infty$) to the means and variances of $f(x, t)$,

$$\alpha = \langle X \rangle, \qquad \beta = \left\langle \tilde{X}^2 \right\rangle. \tag{8.52}$$

For this case we have $\alpha$ and $\beta$ that are independent of $x_0$. Equation (8.43) for the PDF $f(x, t)$ reduces for this case to

$$f(x,t) = \frac{\int f(x_0,t_0)\, dx_0}{\sqrt{2\pi\left\langle \tilde{X}^2 \right\rangle}} \exp\left\{ -\frac{\left(x - \langle X \rangle\right)^2}{2\left\langle \tilde{X}^2 \right\rangle} \right\} = \frac{1}{\sqrt{2\pi\left\langle \tilde{X}^2 \right\rangle}} \exp\left\{ -\frac{\left(x - \langle X \rangle\right)^2}{2\left\langle \tilde{X}^2 \right\rangle} \right\}, \tag{8.53}$$

where the normalization condition for $f(x_0, t_0)$ is applied. Hence, independent of the initial PDF the PDF $f(x, t)$ does relax asymptotically to a normal PDF.

## 8.4 Stochastic Differential Equations

One way of modeling the evolution of stochastic processes was considered in the previous two sections where equations for the PDF of random variables were introduced. It was argued that the reduction of the Kramers-Moyal equation (8.8) to the Fokker-Planck equation (8.21) represents (at least under many conditions) the most suitable way of constructing a PDF transport equation. An alternative is to postulate differential equations for the evolution of stochastic processes. These differential equations determine all the coefficients in the Kramers-Moyal equation. Therefore, this approach does result, too, in a PDF transport equation. The relations between these two approaches will be considered in this section.

### *8.4.1 Nonlinear Markovian Stochastic Equations*

**Approach.** The structure of appropriate stochastic differential equations will be determined in the following way:

- First, the stochastic difference equations developed in Chap. 6 will be used to determine the general structure of stochastic difference equations.
- Second, the general stochastic difference equation obtained in the first step will be represented as a stochastic differential equation.
- Third, the stochastic integration will be defined, which is required for the calculation of solutions of stochastic differential equations (and the reproduction of the stochastic differential equation considered, see below).

**Stochastic Difference Equation.** The first step requires the generalization of the stochastic difference equation models that have been developed in Chap. 6. All the models considered can be covered by the equation

$$\frac{X_n - X_{n-1}}{\Delta t} = a(X_{n-1}, t_{n-1}) + b(X_{n-1}, t_{n-1}) \frac{\Delta W_{n-1}}{\Delta t}. \tag{8.54}$$

Here, $X_n$ represents the variable considered (the particle position, particle velocity, or population density), and we have $\Delta W_{n-1} = (\Delta t)^{1/2}\ \varepsilon_{n-1}$. The random variable $\varepsilon_{n-1}$ is normality distributed and characterized by $<\varepsilon_k> = 0$ and $<\varepsilon_k\ \varepsilon_m> = \delta_{km}$. We will assume that $\varepsilon_{n-1}$ is independent of $X_0$. The coefficients $a(X_{n-1}, t_{n-1})$ and $b(X_{n-1}, t_{n-1})$ are any functions of $X_{n-1}$ and $t_{n-1}$ (the models considered in Chap. 6 did not involve an explicit time dependence, but such a $t_{n-1}$ dependence may be relevant to other problems). No assumption about the time interval $\Delta t$ is made here. Regarding the following explanations it is worth noting that $X_{n-1}$ is independent of $\varepsilon_{n-1}$ because $X_{n-1}$ is only affected by previous values $\varepsilon_{n-2}$, $\varepsilon_{n-3}$, .... Hence, $\Delta W_{n-1}$ is independent of the coefficients $a(X_{n-1}, t_{n-1})$ and $b(X_{n-1}, t_{n-1})$.

**Stochastic Differential Equation.** The second step is the transition from the stochastic difference equation (8.54) to a differential equation. By considering an infinitesimal time interval $\Delta t \to 0$ and defining time $t$ by $t = n\ \Delta t$, the stochastic model (8.54) can be written as

$$\frac{dX}{dt}(t) = a(X,t) + b(X,t)\frac{dW}{dt}(t). \tag{8.55}$$

The change of the stochastic process $X$ is fully determined by $a(X(t), t)$, $b(X(t), t)$, and $dW/dt(t)$. Therefore, we find that Eq. (8.55) describes the evolution of the stochastic process $X$ as a Markov process: the future properties of $X$ are fully determined by the present state at $t$. Let us have a closer look at the properties of the derivative $dW/dt$ of a Wiener process. The limit $\Delta t \to 0$ does not change the statistical properties, i.e., $dW/dt$ is normally distributed as $\Delta W_{n-1}/\Delta t$. The mean of $dW/dt$ can be derived by means of the mean of $\Delta W_{n-1}/\Delta t$,

$$\left\langle \frac{\Delta W_{n-1}}{\Delta t} \right\rangle = 0. \tag{8.56}$$

Correspondingly, the mean of $dW/dt$ is given by

$$\left\langle \frac{dW}{dt}(t) \right\rangle = 0. \tag{8.57}$$

The properties of correlations of $dW/dt$ are determined by the relation

$$\left\langle \frac{\Delta W_n}{\Delta t} \frac{\Delta W_m}{\Delta t} \right\rangle \Delta t = \delta_{nm}. \tag{8.58}$$

For a continuous time $t = n\,\Delta t$ and $t' = \mathrm{m}\,\Delta t$, Eq. (8.58) can be written

$$\left\langle \frac{dW}{dt}(t)\frac{dW}{dt}(t') \right\rangle dt = d\theta(t-t'). \tag{8.59}$$

The differential $d\theta$ of the theta function is zero for negative or positive arguments $t - t'$ but not for $t - t' = 0$: at this value the theta function jumps from zero to one such that $d\theta = 1$. Thus, Eq. (8.59) agrees with the consequences of Eq. (8.58). In terms of the definition of the delta function we can write

$$\left\langle \frac{dW}{dt}(t)\frac{dW}{dt}(t') \right\rangle = \delta(t-t'). \tag{8.60}$$

**Stochastic Integration.** The third step is to define the stochastic integration, such that solutions of the stochastic differential equation (8.55) can be calculated. This question will be addressed by asking under which conditions the integration recovers the stochastic differential equation (8.55) considered. For doing this we integrate Eq. (8.55) from $t$ to $t + dt$,

$$X(t+dt) - X(t) = \int_t^{t+dt} a\big(X(s),s\big)\,ds + \int_t^{t+dt} b\big(X(s),s\big)\frac{dW}{ds}(s)\,ds, \tag{8.61}$$

where $dt \to 0$ is an infinitesimal time interval. The integration can be defined in several ways. One way, which is called the Itô definition, is to take the function values of $a(X(s), s)$ and $b(X(s), s)$ at $X(t)$ and $t$, such that

$$X(t+dt) - X(t) = a\big(X(t),t\big)\,dt + b\big(X(t),t\big)\,dW(t). \tag{8.62}$$

Here, $dW(t)$ is given by $dW(t) = W(t + dt) - W(t)$. Another way of defining the integration, which is the Stratonovich definition, is to take $a(X(s), s)$ and $b(X(s), s)$ at the mean value $[X(t + dt) + X(t)] / 2$ and $t$. For non-random variables both definitions of the integration provide the same result for continuous functions. However, this is not the case for random variables because $dW$ scales with $(\Delta t)^{1/2}$: the Stratonovich definition results in a stochastic differential equation that involves a deterministic drift term in addition to the term $a(X(t), t)$ (see Gardiner 1983 and Risken 1984). In the following we will apply the Itô definition of the stochastic integration because the implied Eq. (8.62) recovers the structure of Eq. (8.55).

### *8.4.2 Relationship to the Fokker-Planck Equation*

Next, let us compare the implications of the stochastic model (8.55) with the consequences of the Fokker-Planck equation (8.21). The comparison of stochastic processes defined in different ways requires, first, the comparison of one-point

statistics, and, second, the comparison of correlations. Therefore, we will calculate the evolution of the PDF and correlations implied by the stochastic model (8.55) and compare these results with the implications of the Fokker-Planck equation.

**PDF Equation.** To find the PDF equation that is implied by the stochastic model (8.55) we consider the Kramers-Moyal equation (8.8),

$$\frac{\partial f(x,t)}{\partial t} = \sum_{n=1}^{\infty} \left( -\frac{\partial}{\partial x} \right)^n D^{(n)}(x,t)\, f(x,t). \tag{8.63}$$

The Kramers-Moyal coefficients are defined by Eq. (8.12),

$$D^{(n)}(x,t) = \lim_{\Delta t \to 0} \frac{\left\langle \big(X(t+\Delta t) - X(t)\big)^n \mid X(t) = x \right\rangle}{n!\,\Delta t}. \tag{8.64}$$

According to Eq. (8.62), the stochastic differential equation (8.55) provides for the change $\Delta X(t) = X(t + \Delta t) - X(t)$ the expression

$$X(t+\Delta t) - X(t) = a\big(X(t),t\big)\Delta t + b\big(X(t),t\big)\,\Delta W(t). \tag{8.65}$$

In terms of this expression we find for the Kramers-Moyal coefficients

$$\begin{aligned} D^{(n)}(x,t) &= \lim_{\Delta t \to 0} \frac{\left\langle \big[a\big(X(t),t\big)\Delta t + b\big(X(t),t\big)\,\Delta W(t)\big]^n \mid X(t) = x \right\rangle}{n!\,\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{(\Delta t)^{n/2}}{n!\,\Delta t} \left\langle \left[ a\big(x,t\big)\sqrt{\Delta t} + b\big(x,t\big)\frac{\Delta W(t)}{\sqrt{\Delta t}} \right]^n \right\rangle. \end{aligned} \tag{8.66}$$

The second line makes use of $X(t) = x$. The consideration of the factor $(\Delta t)^{n/2}$ simplifies the calculations of Kramers-Moyal coefficients, for which we find

$$D^{(1)}(x,t) = \lim_{\Delta t \to 0} \frac{(\Delta t)^{1/2}}{\Delta t} \left\langle a\big(x,t\big)\sqrt{\Delta t} + b\big(x,t\big)\frac{\Delta W(t)}{\sqrt{\Delta t}} \right\rangle = a\big(x,t\big), \tag{8.67a}$$

$$\begin{aligned} D^{(2)}(x,t) &= \lim_{\Delta t \to 0} \frac{\Delta t}{2\Delta t} \left\langle a^2\big(x,t\big)\Delta t + 2\,a\big(x,t\big)b\big(x,t\big)\,\Delta W(t) + b^2\big(x,t\big)\frac{[\Delta W(t)]^2}{\Delta t} \right\rangle \\ &= \frac{1}{2} b^2\big(x,t\big). \end{aligned} \tag{8.67b}$$

$$D^{(3)}(x,t) = D^{(4)}(x,t) = \cdots D^{(\infty)}(x,t) = 0. \tag{8.67c}$$

The first two expressions result from the properties of $\Delta W(t)$ and the limit $\Delta t \to 0$. The validity of Eq. (8.67c) can be seen by having a closer look at Eq. (8.66). The bracket term provides for all $n$ any finite value. The ratio $(\Delta t)^{n/2} / \Delta t$ is always zero for a vanishing $\Delta t$ if $n \geq 3$. Hence, all Kramers-Moyal coefficients have to vanish

except $D^{(1)}$ and $D^{(2)}$. Thus, the PDF of the stochastic model (8.55) is determined by a Fokker-Planck equation with $D^{(1)} = a$ and $D^{(2)} = b^2/2$. Hence, by using $D^{(1)} = a$ and $D^{(2)} = b^2/2$ we can apply Eqs. (8.25) and (8.32) for the mean $<X>$ and variance $<\widetilde{X}^2>$ implied by the Fokker-Planck equation (8.21) to obtain the equations for the mean and variance that follow from the stochastic model (8.55).

**Correlations.** Next, we will use the stochastic model (8.55) for the calculations of correlations between $X(t)$ and $X(t')$. Consistent with the approach in Sect. 8.2.3 we assume again that $t \le t' = t + r$, where $r$ is any non-negative time. We may consider Eq. (8.55) at $t + r$ instead at $t$,

$$\frac{dX(t+r)}{d(t+r)} = a\big(X(t+r),t+r\big) + b\big(X(t+r),t+r\big)\frac{dW}{dt}(t+r). \tag{8.68}$$

The differentiation of $X$ by $t + r$ can be replaced by a derivative by $r$. We multiply this equation with $\widetilde{X}(t)$ and average,

$$\begin{aligned}\left\langle \widetilde{X}(t)\frac{dX(t+r)}{dr}\right\rangle &= \left\langle \widetilde{X}(t)\,a\big(X(t+r),t+r\big)\right\rangle \\ &\quad + \left\langle \widetilde{X}(t)\,b\big(X(t+r),t+r\big)\frac{dW}{dt}(t+r)\right\rangle.\end{aligned} \tag{8.69}$$

The last term disappears: $dW/dt$ at $t + r$ is independent of $X(t)$ and $X(t + r)$, and $dW/dt$ vanishes in the mean,

$$\left\langle \widetilde{X}(t)\,b\big(X(t+r),t+r\big)\frac{dW}{dt}(t+r)\right\rangle = \left\langle \widetilde{X}(t)\,b\big(X(t+r),t+r\big)\right\rangle\left\langle \frac{dW}{dt}(t+r)\right\rangle = 0. \tag{8.70}$$

By writing the derivative by $r$ in front of the bracket term we obtain, therefore,

$$\frac{d\left\langle \widetilde{X}(t)X(t+r)\right\rangle}{dr} = \left\langle \widetilde{X}(t)\,a\big(X(t+r),t+r\big)\right\rangle. \tag{8.71}$$

We may replace $X$ and $a$ by the corresponding fluctuations because the mean of $X$ and $a$ do not change the correlations functions. Thus, we obtain the equation

$$\frac{d\left\langle \widetilde{X}(t)\widetilde{X}(t+r)\right\rangle}{dr} = \left\langle \widetilde{X}(t)\,\widetilde{a}\big(X(t+r),t+r\big)\right\rangle. \tag{8.72}$$

This equation recovers Eq. (8.33), which is a consequence of the Fokker-Planck equation (8.21) if $D^{(1)} = a$ is taken into account. In combination with the observation that both the stochastic model (8.55) and the Fokker-Planck equation (8.21) imply the same PDF evolution equation, we find that the stochastic model and the Fokker-Planck equation describe the same stochastic process (if the $D^{(1)} = a$ and $D^{(2)} = b^2/2$ apply).

### 8.4.3 Linear Markovian Stochastic Equations

Let us consider an example to make the analysis of nonlinear stochastic models in the previous sections more explicit. We consider the following linear stochastic Markovian differential equation,

$$\frac{dX}{dt}(t) = -\frac{X - \langle X \rangle}{\tau} + \frac{\sqrt{D}}{\tau}\frac{dW}{dt}(t), \tag{8.73}$$

which is relevant regarding the discussion in Sect. 8.5. This equation is written according to the Brownian motion velocity model (6.59b). Here, the constant $D$ is a diffusion coefficient, and the constant $\tau$ is the characteristic relaxation time scale of fluctuations. By averaging Eq. (8.73) the mean value $<X>$ becomes

$$\frac{d\langle X \rangle}{dt} = 0. \tag{8.74}$$

Hence, the mean $<X>$ has to be a constant, this means $<X> = <X_0>$.

**PDF.** The combination of the Fokker-Planck equation (8.21) with the relations $D^{(1)} = a = -(x - <X>) / \tau$ and $D^{(2)} = b^2 / 2 = D / (2\,\tau^2)$ between the coefficients of the PDF equation (8.21) and the stochastic model (8.55) shows that the evolution of $X(t)$ is described by the Fokker-Planck equation

$$\frac{\partial f(x,t)}{\partial t} = \frac{1}{\tau}\frac{\partial \left(x - \langle X \rangle\right) f(x,t)}{\partial x} + \frac{D}{2\tau^2}\frac{\partial^2 f(x,t)}{\partial x^2}. \tag{8.75}$$

This equation represents a specific case of the PDF equation (8.34), which can be solved analytically. In particular, we have here the case that $F = 0$, $G = -1 / \tau$, and $D$ in Eq. (8.34) corresponds to $D / (2\,\tau^2)$ in Eq. (8.75). Hence, the solution $f(x, t)$ of Eq. (8.75) is given by Eq. (8.43),

$$f(x,t) = \int \frac{1}{\sqrt{2\pi\beta}} \exp\left\{-\frac{(x-\alpha)^2}{2\beta}\right\} f(x_0,t_0)\, dx_0. \tag{8.76}$$

Here, $f(x_0, t_0)$ refers to any initial PDF. The model parameters $\alpha$ and $\beta$ satisfy the Eqs. (8.48),

$$\frac{d\alpha}{dt} = -\frac{\alpha - \langle X \rangle}{\tau}, \tag{8.77a}$$

$$\frac{d\beta}{dt} = -\frac{2\beta}{\tau} + \frac{D}{\tau^2}. \tag{8.77b}$$

The initial conditions for these equations are given by $\alpha(0) = x_0$ and $\beta(0) = 0$. For simplicity, we assume here that $t_0 = 0$. The solutions of Eqs. (8.77) become

very simple by writing these equations in the following way,

$$\frac{d}{dt}\left(\alpha - \langle X\rangle\right) = -\frac{1}{\tau}\left(\alpha - \langle X\rangle\right), \tag{8.78a}$$

$$\frac{d}{dt}\left(\beta - \frac{D}{2\tau}\right) = -\frac{2}{\tau}\left(\beta - \frac{D}{2\tau}\right). \tag{8.78b}$$

Correspondingly, the model parameters $\alpha$ and $\beta$ are given by the functions

$$\alpha = \langle X\rangle + \left(x_0 - \langle X\rangle\right)e^{-\frac{t}{\tau}}, \tag{8.79a}$$

$$\beta = \frac{D}{2\tau} - \frac{D}{2\tau}e^{-\frac{2t}{\tau}} = \frac{D}{2\tau}\left(1 - e^{-\frac{2t}{\tau}}\right). \tag{8.79b}$$

Asymptotically, $\alpha$ realizes <$X$> and $\beta$ realizes $D/(2\tau)$.

**Correlations.** For the calculation of correlations of $X$ it is helpful to consider the formal solution of Eq. (8.73), which is given by

$$X(t) = \langle X\rangle + \left(X_0 - \langle X\rangle\right)e^{-\frac{t}{\tau}} + \frac{\sqrt{D}}{\tau}\int_0^t e^{-\frac{t-s}{\tau}}\frac{dW}{ds}(s)\,ds. \tag{8.80}$$

By setting $t = 0$ we can see that this solution recovers the initial condition $X_0$. The fact that $X(t)$ satisfies the differential equation (8.73) can be seen by taking the derivative of this expression. Therefore, the fluctuations $\widetilde{X}(t) = X(t) - \langle X\rangle$ of $X$ are given by

$$\widetilde{X}(t) = \widetilde{X}_0\, e^{-\frac{t}{\tau}} + \frac{\sqrt{D}}{\tau}\int_0^t e^{-\frac{t-s}{\tau}}\frac{dW}{ds}(s)\,ds. \tag{8.81}$$

To calculate the correlation function we multiply $\widetilde{X}(t)$ with

$$\widetilde{X}(t') = \widetilde{X}_0\, e^{-\frac{t'}{\tau}} + \frac{\sqrt{D}}{\tau}\int_0^{t'} e^{-\frac{t'-s'}{\tau}}\frac{dW}{ds'}(s')\,ds', \tag{8.82}$$

where $t'$ is any time. By taking the average of this product we find

$$\begin{aligned}\left\langle \widetilde{X}(t)\widetilde{X}(t')\right\rangle &= \\ &= \left\langle \left(\widetilde{X}_0\, e^{-\frac{t}{\tau}} + \frac{\sqrt{D}}{\tau}\int_0^t e^{-\frac{t-s}{\tau}}\frac{dW}{ds}(s)\,ds\right)\left(\widetilde{X}_0\, e^{-\frac{t'}{\tau}} + \frac{\sqrt{D}}{\tau}\int_0^{t'} e^{-\frac{t'-s'}{\tau}}\frac{dW}{ds'}(s')\,ds'\right)\right\rangle \\ &= \left\langle \widetilde{X}_0{}^2\right\rangle e^{-\frac{t+t'}{\tau}} + \frac{D}{\tau^2}\int_0^t\int_0^{t'} e^{-\frac{t-s}{\tau}}e^{-\frac{t'-s'}{\tau}}\left\langle \frac{dW}{ds}(s)\frac{dW}{ds'}(s')\right\rangle ds'\,ds. \end{aligned} \tag{8.83}$$

In accordance with the fact that $\varepsilon_{n-1}$ is independent of $X_0$ (see the discussion of Eq. (8.54)) we find that $dW/ds$ and $dW/ds'$ are uncorrelated to $\widetilde{X}_0$, which leads to the last expression. The last term in Eq. (8.83) can be rewritten,

$$\begin{aligned}
\frac{D}{\tau^2}\int_0^t\int_0^{t'} e^{-\frac{t-s}{\tau}} e^{-\frac{t'-s'}{\tau}} \left\langle \frac{dW}{ds}(s)\frac{dW}{ds'}(s')\right\rangle ds'\,ds &= \frac{D}{\tau^2}\int_0^t\int_0^{t'} e^{-\frac{t-s}{\tau}} e^{-\frac{t'-s'}{\tau}} \delta(s'-s)\,ds'\,ds \\
&= \frac{D}{\tau^2}\int_0^t\int_0^{t'} e^{-\frac{t+t'-2s}{\tau}} \frac{d\theta(s'-s)}{ds'}\,ds'\,ds = \frac{D}{\tau^2}\int_0^t e^{-\frac{t+t'-2s}{\tau}} \left[\theta(t'-s)-\theta(-s)\right] ds \\
&= \frac{D}{\tau^2}\int_0^t e^{-\frac{t+t'-2s}{\tau}} \theta(t'-s)\,ds = \frac{D}{\tau^2}\int_0^{\min(t,t')} e^{-\frac{t+t'-2s}{\tau}}\,ds \\
&= \frac{D}{\tau^2}\frac{\tau}{2}\left[e^{-\frac{t+t'-2\min(t,t')}{\tau}} - e^{-\frac{t+t'}{\tau}}\right] = \frac{D}{2\tau}\left[e^{-\frac{|t-t'|}{\tau}} - e^{-\frac{t+t'}{\tau}}\right].
\end{aligned} \tag{8.84}$$

The first line applies the definition of the correlation of $dW/ds$ and $dW/ds'$. In the second line, the sifting property of the delta function is used, the delta function is replaced by the derivative of the theta function, and the integration over $ds'$ is performed. Here, the term $\theta(-s)$ is zero due to the variation of $s$ considered. The remaining expressions provide then an explicit function. Hence, Eq. (8.83) reads

$$\begin{aligned}
\left\langle \widetilde{X}(t)\widetilde{X}(t')\right\rangle &= \left\langle \widetilde{X}_0{}^2\right\rangle e^{-\frac{t+t'}{\tau}} + \frac{D}{2\tau}\left[e^{-\frac{|t-t'|}{\tau}} - e^{-\frac{t+t'}{\tau}}\right] \\
&= \frac{D}{2\tau}e^{-\frac{|t-t'|}{\tau}} + \left(\left\langle \widetilde{X}_0{}^2\right\rangle - \frac{D}{2\tau}\right)e^{-\frac{t+t'}{\tau}}.
\end{aligned} \tag{8.85}$$

Without loss of generality we set $t \le t' = t + r$, where $r$ is a non-negative time interval,

$$\left\langle \widetilde{X}(t)\widetilde{X}(t+r)\right\rangle = \frac{D}{2\tau}e^{-\frac{r}{\tau}} + \left[\left\langle \widetilde{X}_0{}^2\right\rangle - \frac{D}{2\tau}\right]e^{-\frac{2t+r}{\tau}} = e^{-\frac{r}{\tau}}\left(\frac{D}{2\tau} + \left[\left\langle \widetilde{X}_0{}^2\right\rangle - \frac{D}{2\tau}\right]e^{-\frac{2t}{\tau}}\right). \tag{8.86}$$

The latter result can be written more conveniently as

$$\left\langle \widetilde{X}(t)\widetilde{X}(t+r)\right\rangle = e^{-\frac{r}{\tau}}\left\langle \widetilde{X}^2(t)\right\rangle, \tag{8.87}$$

where the variance is defined by

$$\left\langle \widetilde{X}^2(t)\right\rangle = \frac{D}{2\tau} + \left(\left\langle \widetilde{X}_0{}^2\right\rangle - \frac{D}{2\tau}\right)e^{-\frac{2t}{\tau}}. \tag{8.88}$$

The last to expressions agree with the corresponding consequences of the Fokker-Planck equation (8.21) (see exercise 8.4.1).

## 8.5 Non-Markovian Stochastic Differential Equations

In the previous sections we developed mathematical concepts for the modeling of the evolution of stochastic processes. We did not address at all the question of how these concepts can be applied to the solution of a problem, this means how a case considered can be modeled in terms of a stochastic differential equation. The stochastic differential equations introduced above are general with one exception: the essential assumption applied was the consideration of a Markovian stochastic process (the assumption that the future statistical properties of a stochastic process are fully determined by the present process properties). However, this assumption is usually not rigorously satisfied: most real processes represent non-Markovian processes. Let us consider how we can deal with this problem by considering a relevant but not too complicated problem: the motion of a molecule.

### *8.5.1 Markovian and Non-Markovian Velocity Models*

**Velocity Model.** Let us consider the following Markovian linear stochastic model for the velocity $v$ of one molecule in one direction (Heinz 2003, 2004),

$$\frac{dv}{dt}(t) = -\frac{v(t) - \langle v \rangle}{\tau} + \sqrt{\frac{4e}{3\tau}} \frac{dW}{dt}. \tag{8.89}$$

Here, $<v>$ is the mean molecular velocity, $\tau$ is the characteristic relaxation time scale of velocities, and $e$ refers to the specific kinetic energy of molecules, which is related to the equilibrium variance $<\tilde{v}_e^{\,2}>$ of the velocity by $e = 3 <\tilde{v}_e^{\,2}> / 2$. By averaging Eq. (8.89) we see that $<v>$ has to be constant. For simplicity, the parameters $\tau$ and $e$ are considered to be constant. This velocity model represents the continuous version of the discrete velocity model (6.59b) for the motion of a Brownian particle, where the diffusion coefficient $D$ is replaced by $D = 4\, e\, \tau / 3$ (the only difference is given by the consideration of a nonzero mean velocity $<v>$ here).

**Acceleration Model.** The velocity model (8.89) represents a reasonable model, but its simplicity implies some shortcomings (see the discussion in Sect. 8.5.4). A model that does better agree with the motion of molecules in reality is given by the following stochastic equation for the acceleration $dv/dt$ (Heinz 2003, 2004),

$$\frac{d}{dt}\frac{dv}{dt}(t) = \frac{1}{\tau_f}\left[ -\frac{dv}{dt} - \frac{v(t) - \langle v \rangle}{\tau} + \sqrt{\frac{4e}{3\tau}} \frac{dW}{dt} \right]. \tag{8.90}$$

The velocity model (8.89) can be recovered by multiplying this equation by $\tau_f$ and taking the limit $\tau_f \to 0$. The bracket term is equal to zero in this case, which corresponds to the velocity model (8.89). The idea of Eq. (8.90) is that there are nonzero deviations from the velocity model assumption (8.89), which are given by the bracket term. The latter deviations lead to accelerations that relax to zero with a characteristic time scale $\tau_f$. For simplicity, $\tau_f$ is considered to be constant as $\langle v \rangle$, $\tau$, and $e$. It is interesting that the model (8.90) also can be written as a system of equations by considering the set of variables ($v$, $a$), where $a = dv/dt$ is the particle acceleration. According to the definition $a = dv/dt$ and Eq. (8.90), the equations for these variables are given by

$$\frac{dv}{dt}(t) = a, \tag{8.91a}$$

$$\frac{da}{dt}(t) = \frac{1}{\tau_f}\left[ -a(t) - \frac{v(t) - \langle v \rangle}{\tau} + \sqrt{\frac{4e}{3\tau}}\,\frac{dW}{dt} \right]. \tag{8.91b}$$

These equations represent a linear Markovian stochastic equation system for the process ($v$, $a$), this means we have a Markovian acceleration model. The fact that $\langle v \rangle$ is constant implies that $\langle a \rangle = 0$.

**Velocity Model Implied by Acceleration Model.** Let us write the acceleration model (8.90) as a velocity model to compare it with the velocity model (8.89). Equation (8.90) can be seen as a nonhomogeneous linear first-order equation for $dv/dt$. The formal solution of this equation reads

$$\frac{dv}{dt} = a_0\, e^{-t/\tau_f} + \frac{1}{\tau_f}\int_0^t e^{-(t-s)/\tau_f}\left[ -\frac{v(s) - \langle v \rangle}{\tau} + \sqrt{\frac{4e}{3\tau}}\,\frac{dW}{ds}(s) \right] ds. \tag{8.92}$$

The validity of this writing may be seen by proving that this solution provides the initial acceleration $a(0) = a_0$, and it satisfies the differential equation (8.90). On the right-hand side we have terms of different nature: the term with $(v(s) - \langle v \rangle)/\tau$ is a systematic contribution, and the terms that involve $a_0$ and $dW/ds$ are random terms. To simplify the writing we combine the last two terms to a stochastic force

$$f(t) = a_0\, e^{-t/\tau_f} + \frac{1}{\tau_f}\sqrt{\frac{4e}{3\tau}}\int_0^t e^{-(t-s)/\tau_f}\,\frac{dW}{ds}(s)\, ds. \tag{8.93}$$

This force vanishes in the mean, $\langle f(t) \rangle = 0$, and its initial value is $f_0 = f(0) = a_0$. The simplest way to see the statistical properties of $f(t)$ is to use Eq. (8.93) for the derivation of the following equation for $f(t)$,

$$\frac{df}{dt} + \frac{1}{\tau_f} f = \frac{1}{\tau_f}\sqrt{\frac{4e}{3\tau}}\,\frac{dW}{dt}, \tag{8.94}$$

which enables the use of the analysis results of the linear stochastic model (8.73). By replacing $D$ and $\tau$ in Eq. (8.73) by $D = 4\,e/(3\,\tau)$ and $\tau_f$, respectively, we can use the correlation (8.85) for finding the following force correlation,

$$\left\langle f(t)f(t')\right\rangle = \frac{2e}{3\tau\tau_f}e^{-\frac{|t-t'|}{\tau_f}} + \left(\left\langle {f_0}^2\right\rangle - \frac{2e}{3\tau\tau_f}\right)e^{-\frac{t+t'}{\tau_f}}. \tag{8.95}$$

The setting of the initial force variance is up to us: we set $<f_0^2> = 2\,e/[3\,\tau\tau_f]$ such that the last term in the force correlation, which vanishes asymptotically, is equal to zero. Equation (8.95) shows that the setting of $<f_0^2>$ implies that $<f^2(t)> = <f_0^2>$, this means we thus assume equilibrium conditions. Due to $f_0 = a_0$ we specify in this way $<a_0^2> = 2\,e/[3\,\tau\tau_f]$. Correspondingly, the force $f(t)$ is characterized by

$$\left\langle f(t)\right\rangle = 0, \tag{8.96a}$$

$$\left\langle f(t)f(t')\right\rangle = \frac{2e}{3\tau\tau_f}e^{-\frac{|t-t'|}{\tau_f}} = \frac{\left\langle {\tilde{v}_e}^2\right\rangle}{\tau\tau_f}e^{-\frac{|t-t'|}{\tau_f}}, \tag{8.96b}$$

where the relation $<\tilde{v}_e^{\,2}> = 2\,e/3$ is used. In terms of the definition (8.93) of $f(t)$ and the properties (8.96) we can write the velocity model (8.92) implied by the acceleration model as

$$\frac{dv}{dt} = -\left\langle {\tilde{v}_e}^2\right\rangle^{-1}\int_0^t \left\langle f(t)f(s)\right\rangle\left(v(s) - \left\langle v\right\rangle\right)ds + f(t). \tag{8.97}$$

We see that the force $f(t)$ does control both the generation of fluctuations and the relaxation. This equation represents a non-Markovian velocity model because the future velocity is calculated in terms of the velocity history between zero and $t$.

**Model Comparison.** What is the difference between the Markovian velocity model (8.89) and non-Markovian velocity model (8.97) with regard to the process statistics provided by these models? The mean velocity is constant in both models. Consistent with the consideration of stochastic forces we consider equilibrium conditions. For this case, the velocity variance is independent of $t$, this means we have $<\tilde{v}^2(t)> = <\tilde{v}_0^{\,2}> = <\tilde{v}_e^{\,2}> = 2\,e/3$ in both models. Therefore, the difference between both velocity models will be given, first of all, by the velocity correlation function $<\tilde{v}(t)\,\tilde{v}(t')>$. This correlation function will be calculated and compared in the following sections. The difference between the two models also can be seen regarding the acceleration correlation. It is interesting that the acceleration correlation function is determined by the velocity correlation for the equilibrium conditions considered. To show this we consider

$$\frac{d}{dt}\frac{d}{dr}\left\langle\tilde{v}(t)\,\tilde{v}(t+r)\right\rangle = \frac{d}{dt}\left\langle\tilde{v}(t)\,\tilde{a}(t+r)\right\rangle = \left\langle\tilde{a}(t)\,\tilde{a}(t+r)\right\rangle + \left\langle\tilde{v}(t)\frac{d\tilde{a}(t+r)}{dt}\right\rangle. \tag{8.98}$$

Here, $r$ can be a positive or negative time interval. The left-hand side equals zero under equilibrium conditions because all statistics are independent of $t$ (this means constant). By writing in the last term the derivative by $t$ as a derivative by $r$ we find the condition

$$\left\langle \widetilde{a}(t)\,\widetilde{a}(t+r)\right\rangle = -\frac{d^2}{dr^2}\left\langle \widetilde{v}(t)\,\widetilde{v}(t+r)\right\rangle. \tag{8.99}$$

This finding enables the calculation of the acceleration correlation for any given velocity correlation.

## *8.5.2 Markovian Velocity Model Analysis*

**Velocity Correlation.** The statistical properties of Eq. (8.89) can be obtained by using the results obtained for the linear stochastic model (8.73). By replacing $D$ in Eq. (8.85) by $D = 4\,e\,\tau/3$ we find ($r$ can be positive or negative)

$$\left\langle \widetilde{v}(t)\,\widetilde{v}(t+r)\right\rangle = \frac{2e}{3}e^{-\frac{|r|}{\tau}} + \left(\left\langle \widetilde{v}_0{}^2\right\rangle - \frac{2e}{3}\right)e^{-\frac{2t+r}{\tau}}. \tag{8.100}$$

The last term has to be equal to zero for the equilibrium conditions considered. Thus, we set $<\widetilde{v}_0{}^2> = 2\,e/3$. The use of this setting in Eq. (8.100) then shows that $<\widetilde{v}^2(t)> = <\widetilde{v}_0{}^2>$, which corresponds to the equilibrium considered. The normalized correlation function

$$C_v(t,r) = \frac{\left\langle \widetilde{v}(t)\,\widetilde{v}(t+r)\right\rangle}{\left\langle \widetilde{v}^2(t)\right\rangle} \tag{8.101}$$

is then given by the expression

$$C_v(t,r) = e^{-\frac{|r|}{\tau}}. \tag{8.102}$$

It is interesting that the normalized correlation function of an equilibrium process represents a correlation coefficient. For example, for our case we have

$$\frac{\left\langle \widetilde{v}(t)\,\widetilde{v}(t+r)\right\rangle}{\sqrt{\left\langle \widetilde{v}^2(t)\right\rangle\left\langle \widetilde{v}^2(t+r)\right\rangle}} = \frac{\left\langle \widetilde{v}(t)\,\widetilde{v}(t+r)\right\rangle}{\sqrt{\left\langle \widetilde{v}^2(t)\right\rangle\left\langle \widetilde{v}^2(t)\right\rangle}} = \frac{\left\langle \widetilde{v}(t)\,\widetilde{v}(t+r)\right\rangle}{\left\langle \widetilde{v}^2(t)\right\rangle} = C_v(t,r). \tag{8.103}$$

Therefore, there is the requirement that $|C_v(t, r)| \leq 1$. The normalized correlation function (8.102) shows that the condition $|C_v(t, r)| \leq 1$ is satisfied. The function $C_v(t, r)$ can be used to calculate a time scale $T_v$ that characterizes the correlation

time. This integral time scale $T_v$ is defined by the integral

$$T_v = \int_0^\infty C_v(t,r)dr. \tag{8.104}$$

According to Eq. (8.102), $T_v$ is given by

$$T_v = \int_0^\infty e^{-\frac{r}{\tau}} dr = -\tau\left(\exp\left\{-\frac{r}{\tau}\right\}\right)_0^\infty = \tau. \tag{8.105}$$

This relation shows that $\tau$ determines the correlation time of velocity fluctuations.

**Acceleration Correlation.** The acceleration correlation implied by the velocity model can be calculated by applying Eq. (8.99). We obtain

$$\left\langle \tilde{a}(t)\tilde{a}(t+r)\right\rangle = -\frac{d^2}{dr^2}\left\langle \tilde{v}(t)\,\tilde{v}(t+r)\right\rangle = -\frac{2e}{3}\frac{d^2}{dr^2}e^{-\frac{|r|}{\tau}} = -\frac{2e}{3\tau^2}e^{-\frac{|r|}{\tau}}. \tag{8.106}$$

This consequence of the Markovian velocity model leads to questions, e.g., about the negative acceleration variance $-2e/[3\tau^2]$ predicted by Eq. (8.106). The result (8.106) is implied by Eq. (8.99), which requires the assumption $r \neq 0$. To see the validity of Eq. (8.106) for all values of $r$ we have to calculate the acceleration correlation directly from the Markovian velocity model (8.89). As shown in exercise 8.5.1, this leads to the result

$$\left\langle \tilde{a}(t)\tilde{a}(t+r)\right\rangle = -\frac{2e}{3\tau^2}e^{-\frac{|r|}{\tau}} + \frac{4e}{3\tau}\delta(r). \tag{8.107}$$

Evidence for the need to involve the delta function in addition to the exponential function provided by the correlation function (8.106) can be obtained by proving that Eq. (8.107) implies

$$\int_0^\infty \left\langle \tilde{a}(t)\tilde{a}(t+r)\right\rangle dr = 0, \tag{8.108}$$

which is a property of any variable (which is here the acceleration) that represents the derivative of an equilibrium process (which is here the velocity): see exercise 8.5.2. This property means that accelerations have a zero integral correlation time. The difference between the correct acceleration correlation (8.107) and the incorrect acceleration correlation (8.106) shows that the use of Eq. (8.99) is problematic if the velocity correlation is not a smooth function of $r$, as given for the model (8.102) at $r = 0$. The acceleration correlation function (8.107) does not lead to a finite acceleration variance at $r = 0$. Hence, a normalized acceleration correlation function and a related integral time scale cannot be calculated.

### 8.5.3 Non-Markovian Velocity Model Analysis

**Velocity Correlation.** The velocity correlation implied by the non-Markovian velocity model is calculated in the last paragraph (called appendix) of this section. For the equilibrium considered we obtain according to Eq. (8.128) the result

$$\left\langle \tilde{v}(t)\tilde{v}(t+r)\right\rangle = \frac{2e}{3}\,\frac{r_1 e^{r_2|r|} - r_2 e^{r_1|r|}}{r_1 - r_2}. \tag{8.109}$$

The parameters $r_1$ and $r_2$ are given by the expressions

$$r_1 = \frac{1}{2\tau_f}\left[-1+\sqrt{1-4\frac{\tau_f}{\tau}}\right], \qquad r_2 = \frac{1}{2\tau_f}\left[-1-\sqrt{1-4\frac{\tau_f}{\tau}}\right]. \tag{8.110}$$

These parameters have the properties

$$r_1 + r_2 = -\frac{1}{\tau_f}, \qquad r_1\, r_2 = \frac{1}{4{\tau_f}^2}\left[1-\left(1-4\frac{\tau_f}{\tau}\right)\right] = \frac{1}{\tau\,\tau_f}, \tag{8.111}$$

which we will use below. The setting $r = 0$ in Eq. (8.109) reveals $<\tilde{v}^2(t)> = 2e/3$. Thus, the normalized velocity correlation function (8.101) reads for this case

$$C_v(t,r) = \frac{r_1 e^{r_2|r|} - r_2 e^{r_1|r|}}{r_1 - r_2}. \tag{8.112}$$

The integral time scale $T_v$ provides for this velocity correlation function

$$\begin{aligned} T_v &= \int_0^\infty C_v(t,r)dr = \int_0^\infty \frac{r_1 e^{r_2 r} - r_2 e^{r_1 r}}{r_1 - r_2}dr = \frac{1}{r_1 - r_2}\left(\frac{r_1}{r_2}e^{r_2 r} - \frac{r_2}{r_1}e^{r_1 r}\right)_0^\infty \\ &= \frac{1}{r_1 - r_2}\left(\frac{r_2}{r_1} - \frac{r_1}{r_2}\right) = \frac{{r_2}^2 - {r_1}^2}{r_1\, r_2 (r_1 - r_2)} = -\frac{r_1 + r_2}{r_1\, r_2} = \tau. \end{aligned} \tag{8.113}$$

The exponential functions become zero at infinity: $r_1$ and $r_2$ are negative because the square root in the definitions (8.110) is smaller than one. The application of Eq. (8.111) implies $T_v = \tau$, which is the same result as obtained for the Markovian velocity model (see Eq. (8.105)).

**Acceleration Correlation.** Based on the velocity correlation we can calculate the acceleration correlation. In terms of Eq. (8.99) we find

$$\begin{aligned} \left\langle \tilde{a}(t)\,\tilde{a}(t+r)\right\rangle &= -\frac{d^2}{dr^2}\left\langle \tilde{v}(t)\tilde{v}(t+r)\right\rangle = -\frac{2e}{3}\,\frac{d^2}{dr^2}\,\frac{r_1 e^{r_2|r|} - r_2 e^{r_1|r|}}{r_1 - r_2} \\ &= -\frac{2e}{3}\,r_1 r_2\,\frac{r_2 e^{r_2|r|} - r_1 e^{r_1|r|}}{r_1 - r_2} = \frac{2e}{3\tau\tau_f}\,\frac{r_2 e^{r_2|r|} - r_1 e^{r_1|r|}}{r_2 - r_1}, \end{aligned} \tag{8.114}$$

where Eq. (8.111) is used for $r_1$ $r_2$. By setting $r = 0$ we get $<\widetilde{a}^2(t)> = 2\,e\,/\,[3\,\tau\,\tau_f]$. In correspondence to the normalized velocity correlation function (8.101) we can define a normalized acceleration correlation function by the expression

$$C_a(t,r) = \frac{\left\langle \widetilde{a}(t)\,\widetilde{a}(t+r)\right\rangle}{\left\langle \widetilde{a}^2(t)\right\rangle}. \tag{8.115}$$

According to Eq. (8.114) and $<\widetilde{a}^2(t)> = 2\,e\,/\,[3\,\tau\,\tau_f]$ we find that

$$C_a(t,r) = \frac{r_2\,e^{r_2|r|} - r_1\,e^{r_1|r|}}{r_2 - r_1}. \tag{8.116}$$

The acceleration integral time scale $T_a$ can be calculated in correspondence to the calculation of the velocity integral time scale $T_v$ by the relation

$$T_a = \int_0^\infty C_a(t,r)\,dr. \tag{8.117}$$

In terms of Eq. (8.116) and the fact that $r_1$ and $r_2$ are negative we find

$$T_a = \int_0^\infty \frac{r_2\,e^{r_2 r} - r_1\,e^{r_1 r}}{r_2 - r_1}\,dr = \frac{1}{r_1 - r_2}\left(e^{r_2 r} - e^{r_1 r}\right)_0^\infty = 0, \tag{8.118}$$

which means that accelerations have a zero integral correlation time.

**Appendix: Velocity Correlation Derivation.** Let us show how the velocity correlation (8.109) implied by the non-Markovian velocity model can be obtained. First, we have to calculate the instantaneous velocity fluctuation that is implied by the acceleration model (8.90). Equation (8.90) represents a linear second-order differential equation that is driven by the time-depending derivative of the Wiener process. By following the solution approach described in Chap. 7 we find

$$\widetilde{v}(t) = \frac{1}{r_2 - r_1}\left[\sqrt{-\frac{4e}{3}\,r_1\,r_2\,(r_1 + r_2)}\int_0^t g(t-s)\,\frac{dW}{ds}(s)\,ds + \widetilde{a}_0\,g(t) - \widetilde{v}_0\,h(t)\right]. \tag{8.119}$$

The parameters $r_1$ and $r_2$, which are defined by Eqs. (8.110), are the roots of the characteristic equation of the linear second-order differential equation. The functions $g(t)$ and $h(t)$ in Eq. (8.119) are abbreviations that are given by

$$g(t) = e^{r_2 t} - e^{r_1 t}, \qquad h(t) = r_1\,e^{r_2 t} - r_2\,e^{r_1 t}. \tag{8.120}$$

The initial acceleration and velocity fluctuations $\widetilde{a}_0$ and $\widetilde{v}_0$ are considered to be uncorrelated. The latter assumption is correct for an equilibrium process because

$$\left\langle \widetilde{a}_0\,\widetilde{v}_0\right\rangle = \left\langle \widetilde{v}(t)\,\frac{d\widetilde{v}}{dt}(t)\right\rangle_{t=0} = \frac{1}{2}\left[\frac{d\left\langle \widetilde{v}^2(t)\right\rangle}{dt}\right]_{t=0} = 0. \tag{8.121}$$

According to Eq. (8.119), the velocity fluctuation at another time $t'$ is given by

$$\tilde{v}(t') = \frac{1}{r_2 - r_1}\left[\sqrt{-\frac{4e}{3} r_1 r_2 (r_1 + r_2)}\int_0^{t'} g(t'-s')\frac{dW}{ds'}(s')\,ds' + \tilde{a}_0\, g(t') - \tilde{v}_0\, h(t')\right]. \tag{8.122}$$

By multiplying Eqs. (8.119) and (8.122) and taking the ensemble average we find

$$\left\langle \tilde{v}(t)\,\tilde{v}(t')\right\rangle = \frac{1}{(r_2 - r_1)^2}\left[-\frac{4e}{3} r_1 r_2 (r_1 + r_2) Q + \left\langle \tilde{a}_0^{\,2}\right\rangle g(t)\,g(t') + \left\langle \tilde{v}_0^{\,2}\right\rangle h(t)\,h(t')\right], \tag{8.123}$$

where the abbreviation $Q$ is given by

$$Q = \int_0^t\int_0^{t'} g(t-s)\,g(t'-s')\left\langle \frac{dW}{ds}(s)\frac{dW}{ds'}(s')\right\rangle ds'\,ds. \tag{8.124}$$

This integral can be evaluated by using the properties of $dW/dt$,

$$\begin{aligned} Q &= \int_0^t\int_0^{t'} g(t-s)\,g(t'-s')\,\delta(s'-s)\,ds'\,ds = \int_0^t g(t-s)\,g(t'-s)\int_0^{t'}\frac{d\theta(s'-s)}{ds'}\,ds'\,ds \\ &= \int_0^t g(t-s)\,g(t'-s)\left[\theta(t'-s) - \theta(-s)\right] ds = \int_0^{\min(t,t')} g(t-s)\,g(t'-s)\,ds. \end{aligned} \tag{8.125}$$

The calculation of this expression leads to the result

$$\begin{aligned} Q &= -\left(\frac{r_1 r_2\, g(t-s)g(t'-s) + h(t-s)h(t'-s)}{2 r_1 r_2 (r_1 + r_2)}\right)_{s=0}^{s=\min(t,t')} \\ &= -\frac{(r_1 - r_2)\,h(|t-t'|) - r_1 r_2\, g(t)\,g(t') - h(t)\,h(t')}{2 r_1 r_2 (r_1 + r_2)}, \end{aligned} \tag{8.126}$$

where the relations $g(0) = 0$ and $h(0) = r_1 - r_2$ are used. In terms of this expression we can write the correlation function (8.123) as

$$\begin{aligned} \left\langle \tilde{v}(t)\,\tilde{v}(t')\right\rangle = \frac{1}{(r_2 - r_1)^2}\Bigg[&(r_1 - r_2)\frac{2e}{3}h(|t-t'|) - r_1 r_2\frac{2e}{3}g(t)\,g(t') - \frac{2e}{3}h(t)\,h(t') \\ &+ \left\langle \tilde{a}_0^{\,2}\right\rangle g(t)\,g(t') + \left\langle \tilde{v}_0^{\,2}\right\rangle h(t)\,h(t')\Bigg]. \end{aligned} \tag{8.127}$$

Under the equilibrium conditions considered the last four terms cancel because of $\langle \tilde{v}_0^{\,2}\rangle = 2e/3$ and $\langle a_0^2\rangle = 2e/[3\,\tau\,\tau_f] = 2 r_1 r_2\, e/3$. Correspondingly, the velocity correlation is found to be given by

$$\left\langle \tilde{v}(t)\,\tilde{v}(t')\right\rangle = \frac{2e}{3}\,\frac{r_1 e^{r_2|t-t'|} - r_2 e^{r_1|t-t'|}}{r_1 - r_2}. \tag{8.128}$$

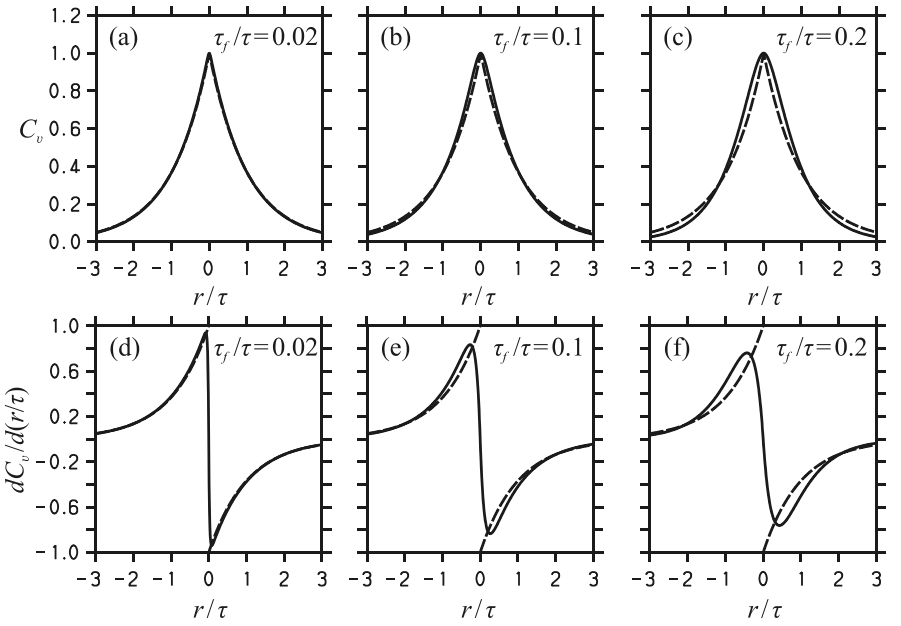



**Fig. 8.2.** The first row shows the normalized velocity correlation function $C_v$ that is implied by the non-Markovian (*solid line*) and Markovian (*dashed line*) velocity model. The velocity time scale $\tau = 1$, and the ratio $\tau_f / \tau$ is given in the plots. The second row shows the derivative of the normalized velocity correlation function for the non-Markovian (*solid line*) and Markovian (*dashed line*) velocity model.

### *8.5.4 The Relevance of Memory Effects*

What is the difference between the Markovian velocity model (8.89) and the non-Markovian velocity model (8.97) regarding the process statistics provided by these models, this means what is the relevance of memory effects? Let us compare the acceleration and velocity correlations of both models to address this question.

**Velocity Correlations.** The velocity correlation functions (8.102) and (8.112) are illustrated in Fig. 8.2. It can be seen that the Markovian and non-Markovian models show a similar behavior: there is only a very minor difference between the curves for the range $\tau_f / \tau \leq 0.2$ considered (we have the condition $\tau_f / \tau \leq 0.25$ due to the definitions of $r_1$ and $r_2$). The area below the curves is equal to $\tau = 1$ for all the cases considered. The effect of the model on the velocity correlation can be better seen by looking at the derivative of the normalized velocity correlations.

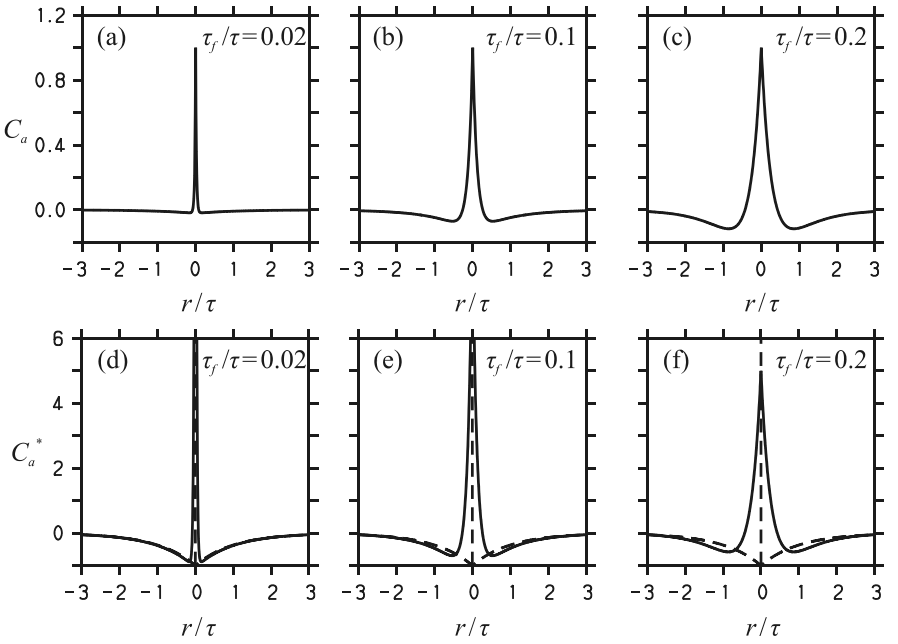



**Fig. 8.3.** The normalized acceleration correlation function $C_a$ implied by the non-Markovian velocity model is shown in the *upper row*. The corresponding function $C_a^*$ is shown in the *lower row* for the non-Markovian velocity model (*solid line*) and the Markovian velocity model (*dashed line*). The velocity time scale $\tau = 1$, and $\tau_f / \tau$ is given in the plots.

Regarding the velocity correlation implied by the non-Markovian and Markovian velocity model, respectively, we find

$$\frac{dC_v}{d(r/\tau)} = r_1 r_2 \tau \frac{r}{|r|} \frac{e^{r_2|r|} - e^{r_1|r|}}{r_1 - r_2}, \qquad \frac{dC_v}{d(r/\tau)} = -\frac{r}{|r|} e^{-\frac{|r|}{\tau}}. \tag{8.129}$$

The corresponding features of the derivatives of normalized velocity correlation functions are also shown in Fig. 8.2. The non-Markovian velocity model implies a smoothly changing velocity correlation, which has a continuous derivative. This behavior is supported by consequences of the Navier-Stokes equations (Sawford 1991, Pope 1994). In contrast, the Markovian velocity model leads to a velocity correlation with a derivative that jumps at $r = 0$. This unphysical behavior is implied by the neglect of acceleration correlations (the neglect of a nonzero $\tau_f$).

**Acceleration Correlations.** The normalized acceleration correlation function $C_a$ that is implied by the non-Markovian velocity model is shown in Fig. 8.3 for different $\tau_f / \tau$, where $\tau = 1$. Such curves cannot be obtained from the Markovian velocity model because the normalized acceleration correlation function cannot be

defined for this case: see the discussion of Eq. (8.107). We see that the condition $|C_a(t, r)| \leq 1$ (see the corresponding discussion regarding $C_v$) for the normalized correlation function of an equilibrium process is satisfied. Due to the fact that the integral acceleration time scale $T_a$ is equal to zero according to Eq. (8.118) we find negative values of the normalized correlation function $C_a$. These features can be compared with the consequences of the Markovian velocity model by consider-ing the following normalized acceleration correlation

$$C_a^*(t,r) = \frac{\langle \tilde{a}(t)\,\tilde{a}(t+r)\rangle}{2e/(3\tau^2)}. \tag{8.130}$$

Here, $\tau_f$ in the definition (8.115) of $C_a$ is replaced by $\tau$. The corresponding curves are shown in Fig. 8.3 for both velocity models. For the non-Markovian velocity model the $C_a^*$ curves are similar to the behavior of $C_a$. It may be seen that $C_a^*$ of the non-Markovian velocity model converges to $C_a^*$ of the Markovian velocity model in the limit $\tau_f \to \infty$. First of all, the difference between the velocity models is given by the acceleration variance $<\tilde{a}^2(t)>$, which is $2\,e/[3\,\tau\,\tau_f]$ and infinity for the non-Markovian and Markovian velocity models, respectively.

**Summary.** These observations can be summarized in the following way: The neglect of memory effects, which means the use of the Markovian velocity model instead of the non-Markovian velocity model, corresponds to the assumption that the characteristic correlation time $\tau_f$ of stochastic forces is negligibly small. This approach is equivalent to the consideration of velocities over time steps that are large compared to $\tau_f$, i.e., the real process is described only asymptotically in this case. The latter approach provides velocity correlations that are very close to the correlations implied by the non-Markovian velocity model. On the other hand, no attempt is made to represent acceleration correlations in a physically correct way: these correlations are only represented such that the integral over the acceleration correlation function is equal to zero, as required for a variable that represents the derivative of an equilibrium process (see the discussion related to Eq. (8.108)).

## 8.6 Summary

The goal of the presentation in this chapter was to develop a methodological basis for the modeling of the evolution of stochastic processes. The latter requires answers to the questions considered in the introduction, this means: the questions about the type of equations for PDFs, their solution, the type of stochastic process equations, and the determination of stochastic equations for the modeling of any case. Let us summarize the features observed.

**PDF Evolution Equation.** Our starting point was the most general equation for the evolution of a PDF given by the Kramers-Moyal equation (8.8),

$$\frac{\partial f(x,t)}{\partial t} = \sum_{n=1}^{\infty} \left( -\frac{\partial}{\partial x} \right)^n D^{(n)}(x,t) f(x,t). \tag{8.131}$$

The Kramers-Moyal equation represents an identity. We did not use any physical principle, we did only assume that the PDF $f(x, t)$ and Kramers-Moyal coefficients $D^{(n)}(x, t)$ exist. The Kramers-Moyal equation implies Pawula's theorem that shows that there are two possibilities: we can either work with an equation that involves an infinite number of Kramers-Moyal coefficients $D^{(n)}(x, t)$, or we can work with the Fokker-Planck equation (8.21),

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial D^{(1)}(x,t) f(x,t)}{\partial x} + \frac{\partial^2 D^{(2)}(x,t) f(x,t)}{\partial x^2}, \tag{8.132}$$

which does only involve the first two Kramers-Moyal coefficients. The neglect of $D^{(n)}$ with $n \geq 3$ is justified if the stochastic process considered has a continuous sample path, this means if jump processes (i.e., processes involving instantaneous unbounded changes) are not considered. We applied the latter assumption in the following. We did also assume that the coefficients $D^{(n)}(x, t)$ do only depend on $x$ and $t$. This corresponds to the assumption that the stochastic process considered is a Markovian process (this means a process for which the present state determines the future evolution).

**Solutions of the Fokker-Planck Equation.** An important question is how we can solve the Fokker-Planck equation (8.132). It was shown that this equation can be solved analytically if we consider the specific Fokker-Planck equation (8.34),

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[F(t) + G(t)\left(x - \langle X \rangle\right)\right] f(x,t) + \frac{\partial^2 D(t) f(x,t)}{\partial x^2}. \tag{8.133}$$

The significant difference between the general Fokker-Planck equations (8.132) and Eq. (8.133) is that $D^{(1)}$ is a linear function of $x$ in Eq. (8.133), and $D^{(2)}$ is independent of $x$. It turns out that the solution to Eq. (8.133) is given by a normal PDF integrated over the initial condition,

$$f(x,t) = \int \frac{1}{\sqrt{2\pi\beta}} \exp\left\{ -\frac{(x-\alpha)^2}{2\beta} \right\} f(x_0, t_0)\, dx_0. \tag{8.134}$$

Here, the model parameter $\alpha$ and $\beta$ are functions of $t$, and $\alpha$ does also depend on $x_0$. Asymptotically (i.e., for $t \to \infty$), $\alpha$ and $\beta$ relax to the mean $<X>$ and variance $< \widetilde{X}^2 >$ of the process considered. Then, the PDF $f(x, t)$ becomes independent of the initial PDF $f(x_0, t_0)$: $f(x, t)$ is then given by a normal PDF with mean $<X>$ and variance $< \widetilde{X}^2 >$.

**Stochastic Process Equations.** Instead of asking how the PDF of a stochastic process evolves, we may ask how the underlying stochastic process evolves in time. In generalization of the stochastic difference equations considered in Chap. 6, we considered the model (8.55) for the evolution of the stochastic process $X(t)$,

$$\frac{dX}{dt}(t) = a\big(X,t\big) + b\big(X,t\big)\frac{dW}{dt}\big(t\big). \tag{8.135}$$

This approach leads to the question of which PDF transport equation is implied by the stochastic model (8.135). To answer this question we calculated the Kramers-Moyal coefficients that are implied by Eq. (8.135), which resulted in

$$D^{(1)}(x,t) = a(x,t), \tag{8.136a}$$

$$D^{(2)}(x,t) = \frac{1}{2}b^2(x,t). \tag{8.136a}$$

$$D^{(3)}(x,t) = D^{(4)}(x,t) = \cdots D^{(\infty)}(x,t) = 0. \tag{8.136c}$$

Hence, the evolution equation for the PDF related to the stochastic model (8.135) is a Fokker-Planck equation with coefficients specified through Eqs. (8.136). By using the coefficient relations we see that the Fokker-Planck equation (8.133) corresponds to the stochastic model

$$\frac{dX}{dt}(t) = F(t) + G(t)\Big(X(t) - \big\langle X\big\rangle\Big) + \sqrt{2D(t)}\,\frac{dW}{dt}. \tag{8.137}$$

Hence, a linear stochastic model has a PDF that is a normal PDF integrated over the initial condition. The most important advantage of stochastic equations is that these equations can be used to represent Fokker-Planck equations that cannot be solved analytically. Such PDF evolution equations can be solved by Monte Carlo simulation, this means the numerical solution of equivalent stochastic equations (see Chap. 6).

**Application to Modeling.** How can we determine stochastic process equations for the modeling of any case? The stochastic differential equation (8.135) can be used for the modeling of any nonlinear processes. On the other hand, Eq. (8.135) describes a Markovian stochastic process, and this assumption is often not rigorously satisfied (most real processes do represent non-Markovian processes). Thus, there is the question about the suitability of modeling a non-Markovian process in terms of a Markovian stochastic differential equation. To address this question we compared in Sect. 8.5 a non-Markovian with a Markovian velocity model: the more accurate non-Markovian velocity model (which may be seen to represent the reality) was used as a reference model to evaluate the performance of the less accurate Markovian velocity model (which represents an approximate model for

the real process). It was shown that the Markovian velocity model is not incorrect but only less complete than the non-Markovian model. The latter model describes processes that take place over the time scale $\tau_f$ (over which accelerations change) and over the time scale $\tau$ (over which velocities change). On the other hand, the Markovian model does only describe processes that take place over $\tau$. The performance of the Markovian model is acceptable regarding the processes that are described: this model provides velocity correlations that are very close to the correlations implied by the non-Markovian velocity model. It is often only possible to model a part of all the processes observed in reality (there are often processes that take place over a variety of time scales, which vary over orders of magnitude). The application of a Markovian model that provides an accurate description for a certain part of these processes and neglects other (smaller-scale) processes does often turn out to be the most convenient choice.

## 8.7 Exercises

**8.2.1** Show the consistency of the Fokker-Planck equation (8.21) by integrating this equation over the sample space from negative to positive infinity.

**8.2.2** Consider the Fokker-Planck equation (8.21).

a) Calculate the asymptotic solution to this equation (this solution has the property $\partial f / \partial t = 0$).

b) Provide an example for $D^{(1)}$ and $D^{(2)}$ that leads to a PDF that approaches zero for $|x| \to \infty$.

c) Provide an example for $D^{(1)}$ and $D^{(2)}$ that leads to a PDF that diverges for positive or negative $x$ with $|x| \to \infty$.

**8.2.3** Consider the Fokker-Planck equation (8.21). Specify this equation (determine the coefficients of the Fokker-Planck equation) so that

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

represents an asymptotic solution. The normal PDF parameters $\mu$ and $\sigma$ are considered to be constant. Hint: you may assume that $D^{(2)}$ is constant.

**8.2.4** Consider Eq. (8.25) for the mean $\langle X \rangle$, which is implied by the Fokker-Planck equation (8.21). Try to solve this equation for the case that $D^{(1)}(x, t) = -a\,x^2$, where $a$ is any constant.

**8.2.5** Consider Eq. (8.32) for the variance $< \widetilde{X}^2 >$, which is a consequence of the Fokker-Planck equation (8.21).

a) Solve this equation for the case that $D^{(1)}(x, t) = -a\,x$ and $D^{(2)}$ is constant. Here, $a$ is any constant.

b) Find the asymptotic variance according to this equation.

c) Explain the relevance of this result regarding the determination of model parameters.

**8.2.6** Consider the Fokker-Planck equation (8.21).

a) Follow the approach used in Sect. 8.2.3 to find the evolution equation for the central moment of third order $< \widetilde{X}^3 >$.

b) Solve this equation for the case that $D^{(1)}(x, t) = -a\,x$ and $D^{(2)}$ is constant. Here, $a$ is any constant.

**8.3.1** Consider the Fokker-Planck equation (8.21).

a) Follow the explanations in Sect. 8.3.1 to show that the Fokker-Planck equation (8.21) applies to the two-point PDF $f(x, t; x', t') = <\delta\,[x - X(t)]\;\delta\,[x' - X(t')]>$,

$$\frac{\partial f(x,t;x',t')}{\partial t'} + \frac{\partial\, D^{(1)}(x',t')\, f(x,t;x',t')}{\partial x'} - \frac{\partial^2 D^{(2)}(x',t')\, f(x,t;x',t')}{\partial x'^2} = 0.$$

b) Apply the definition $f(x, t; x', t') = <\delta\,[x - X(t)]\;\delta\,[x' - X(t')]>$ to show that the correlation function $< \widetilde{X}(t)\widetilde{X}(t') >$ is defined by

$$\left\langle \widetilde{X}(t)\,\widetilde{X}(t') \right\rangle = \int\int \left(x - \left\langle X \right\rangle\right)\left(x' - \left\langle X' \right\rangle\right) f(x,t;x',t')\,dx\,dx'.$$

c) Use the definition of $< \widetilde{X}(t)\widetilde{X}(t+r) >$, where $r$ is any non-negative time, and the Fokker-Planck equation for the two-point PDF $f(x, t; x', t + r)$ to show that $< \widetilde{X}(t)\widetilde{X}(t+r) >$ satisfies the equation

$$\frac{d\left\langle \widetilde{X}(t)\,\widetilde{X}(t+r) \right\rangle}{dr} = \left\langle \widetilde{X}(t)\,\widetilde{D}^{(1)}\left(X(t+r), t+r\right) \right\rangle.$$

**8.3.2** Consider Eq. (8.40) for the conditional PDF $f(x, t \mid x_0, t_0)$. Calculate the asymptotic solution to this equation (which has the property $\partial f / \partial t = 0$). Assume for simplicity that $G$ and $D$ are constant and $F = 0$.

**8.3.3** Consider the model (8.42) for the conditional PDF $f(x, t \mid x_0, t_0)$. Show that the consistency with the initial condition (8.41) requires that $\alpha$ and $\beta$ have the initial values $\alpha(t_0) = x_0$ and $\beta(t_0) = 0$.

**8.3.4** Consider Eqs. (8.48) for $\alpha$ and $\beta$,

$$\frac{d\alpha}{dt} = G(t)\left(\alpha - \left\langle X \right\rangle\right) + F(t), \qquad \frac{d\beta}{dt} = 2\,G(t)\,\beta + 2\,D(t).$$

a) Show that the following expressions are the solutions of these equations and satisfy the initial values $\alpha(t_0) = x_0$ and $\beta(t_0) = 0$.

$$\alpha = x_0 + \int_{t_0}^{t} \exp\left\{ \int_{t_0}^{t-s+t_0} G(r)\, dr \right\} \left( G(s)\left[ x_0 - \langle X \rangle \right] + F(s) \right) ds.$$

$$\beta = 2 \int_{t_0}^{t} \exp\left\{ 2 \int_{t_0}^{t-s+t_0} G(r)\, dr \right\} D(s)\, ds.$$

b) Specify the solutions for the case that $G$ and $D$ are independent of $t$ and $F = 0$.

**8.3.5** Consider the Fokker-Planck equation (8.34) for the case that $D = 0$.

a) Find the function $\beta(t)$ for this case.

b) Find the solution $f(x, t)$ to the Fokker-Planck equation for this case.

c) Specify the solution $f(x, t)$ obtained in b) for the case that the initial PDF is given by $f(x_0, t_0) = \delta(x_0 - \alpha_0)$, where $\alpha_0$ is a given nonrandom value.

d) Interpret the result obtained in c).

**8.3.6** We consider an instantaneous emission of a substance from a point source, this means the emission of a mass $M$ at time zero at a fixed position $H$. The substance diffuses along the $y$ direction. The mean substance concentration is given by $C(y, t) = M f(y, t)$. Here, $f(y, t)$ refers to the PDF for finding a parcel at time $t$ at a position $y$ (see Sect. 6.3.3). The concentration $C$ is described by the diffusion equation ($D$ is a constant diffusion coefficient)

$$\frac{\partial C(y,t)}{\partial t} = D \frac{\partial^2 C(y,t)}{\partial y^2}.$$

a) Specify the initial concentration $C(y_0, 0)$ for this case.

b) Calculate the solution to the diffusion equation based on the solution of the Fokker-Planck equation (8.34).

**8.4.1** Consider the correlation function (8.87) and variance (8.88).

a) Show that Eq. (8.87) agrees with the consequence (8.33) of the general Fokker-Planck equation.

b) Show that Eq. (8.88) agrees with the consequence (8.32) of the general Fokker-Planck equation.

**8.4.2** Consider the case that you are interested to use the linear stochastic model (8.73) for the modeling of a case considered.

a) How is it possible to determine the parameters $D$ and $\tau$ of the model (8.73) in terms of measured statistics?

b) How is it possible to provide evidence for the suitability of modeling a certain case in terms of a linear stochastic model?

**8.4.3** The mixing of species in water or air can be described by the model

$$\frac{d\phi}{dt} = -\frac{1}{\tau}\left(\phi - \langle\phi\rangle\right) + c\frac{dW}{dt}.$$

Here, $\phi$ refers to the instantaneous mass fraction, which is bounded by zero and one according to its definition, i.e., $0 \le \phi \le 1$. Here, $<\phi>$ is the mean value of $\phi$, $\tau$ is a characteristic mixing time scale, $c$ is a parameter, $t$ is time and $dW/dt$ refers to the derivative of a Wiener process.
For simplicity, we assume that $\tau$, $c$, and $<\phi>$ are constants.

a) Use the stochastic mixing model to derive the equation for the variance $<\widetilde{\phi}^2>$. Solve this equation.
b) Use the stochastic mixing model to derive the corresponding equation for the standardized species mass fraction $\Phi = (\phi - <\phi>)/<\widetilde{\phi}^2>^{1/2}$.
c) Use the equation for $\Phi$ to discuss the consequences of applying a zero model parameter $c$. Relate this discussion to the solution of the variance equation.
d) Use the stochastic mixing model to discuss the disadvantage of using a nonzero $c$. Hint: consider the property $0 \le \phi \le 1$ of $\phi$.

**8.4.4** Continue with exercise 8.4.3. The PDF $f(\theta, t)$, which is related to the stochastic model considered, is given by

$$f(\theta,t) = \int f(\theta,t \mid \theta',t') f(\theta',t') d\theta'.$$

Here, $f(\theta, t \mid \theta', t')$ and $f(\theta', t')$ refer to the conditional PDF and initial PDF, respectively.

a) Provide the evolution equation and initial condition for the conditional PDF $f(\theta, t \mid \theta', t')$.
b) Solve this PDF evolution equation. Provide all the model parameters of $f(\theta, t \mid \theta', t')$ as explicit functions of time.
c) Calculate $f(\theta, t)$ for the case that $f(\theta', t') = \delta(\theta' - <\phi>)$.
d) Describe qualitatively the evolution of $f(\theta, t)$ obtained in this way.

**8.4.5** Consider the stochastic population model discussed in Sect. 6.5,

$$\frac{dP}{dt} = P(1-P)\left(\mu + \sigma\frac{dW}{dt}\right).$$

a) Use Eqs. (8.25) and (8.32) to obtain the equations for $<P>$ and $<\widetilde{P}^2>$.
b) Show that the discrete Eqs. (6.94) and (6.96) derived in Chap. 6 imply for $\Delta t \to 0$ the same equations for the mean and variance of $P$.
c) Present the corresponding equation for the PDF $f(p, t)$.
d) Explain why the equation obtained in c) cannot be solved analytically.

**8.5.1** Consider the Markovian velocity model (8.89).

a) Use the velocity model to show that the acceleration correlation function is given for all values of $r$ by the expression

$$\left\langle \tilde{a}(t)\,\tilde{a}(t+r)\right\rangle = -\frac{2e}{3\tau^2}\,e^{-\frac{|r|}{\tau}} + \frac{4e}{3\tau}\,\delta(r).$$

b) Show that the integral of this correlation over $0 \le r < \infty$ is equal to zero.

**8.5.2** Show the validity of

$$\int_0^\infty \left\langle \tilde{a}(t)\,\tilde{a}(t+r)\right\rangle dr = 0$$

for any variable $a$ that represents the derivative of an equilibrium process. Hint: perform the integration over $\tilde{a}(t+r) = d\tilde{v}(t+r)/dr$ directly.

**8.5.3** Consider the velocity correlation function (8.109), which is implied by the non-Markovian velocity model. Calculate the limit $\tau_f \to 0$ of this velocity correlation function to recover the normalized velocity correlation function (8.102) of the Markovian velocity model.

**8.5.4** Consider the acceleration correlation function (8.114) implied by the non-Markovian velocity model. Calculate the limit $\tau_f \to 0$ of this acceleration correlation function to recover the acceleration correlation function (8.107) of the Markovian velocity model.

**8.5.5** The velocity model (8.89) implies the following model for the position $x$ defined by $dx/dt = v$,

$$\frac{dx}{dt} = \langle v\rangle + F(t).$$

Here, $\langle v\rangle$ is constant, and the stochastic force $F(t)$ is defined by

$$F(t) = \tilde{v}(0)\,e^{-t/\tau} + \sqrt{\frac{4e}{3\tau}}\int_0^t e^{-(t-s)/\tau}\,\frac{dW}{ds}(s)\,ds.$$

a) Show that this position model agrees with the velocity model (8.89).

b) Explain why the position model represents a non-Markovian model.

c) Calculate under equilibrium conditions the mean and correlation of the stochastic force $F(t)$. You may follow the explanations in Sect. 8.5.1.

d) Which condition does this non-Markovian position model reduce to a Markovian position model? To provide the answer to this question you have to specify the force $F(t)$. Use the relation $e = 3\,\nu/\tau$ between $e$ and the kinematic viscosity $\nu$ (see Sect. 10.5). Neglect the first term in the $F(t)$ expression, which is justified under equilibrium conditions.

# 9 Deterministic Multivariate Evolution

The discussions of deterministic evolution in Chap. 7 were focused on the modeling of the evolution of one variable (as heat, mass, the position of any body, or a population density). The consideration of such relatively simple problems is helpful for a basic understanding of the structure and the range of applicability of equations for typical problems. However, only a narrow range of problems can be described in this way: the analysis of most real problems requires the consideration of the multivariate evolution of several variables. The latter is required, for example, regarding the interaction of biological species and motions of bodies or fluids in three-dimensional space. To deal with such cases we extend here the concepts used for the modeling of mechanical and population ecology processes in Chap. 7 to the modeling of the joint evolution of several variables. We will continue with the consideration of global properties that change in time but not in space, i.e., partial differential equations that describe the evolution of processes in space will be not considered. The mathematics of models for the evolution of such processes can be formulated in terms of linear and nonlinear systems of coupled ordinary differential equations.

Section 9.1 explains the motivation for developing mathematical models for the multivariate evolution of processes. Section 9.2 prepares the discussions in the following sections by the explanation of techniques for the solution and analysis of coupled systems of ordinary differential equations. Sections 9.3 and 9.4 extend the discussion in Chap. 7. Section 9.3 describes the modeling of basic population ecology processes (the competition for food and predator-prey interactions). The modeling of mechanical motions will be considered in Sect. 9.4, where the pendulum equation used in Chap. 3 will be solved. Section 9.5 illustrates the problem of dealing with the fluid dynamics equations derived in Chap. 10 by considering a simple model for atmospheric motions. Section 9.6 summarizes the basic features of the modeling approaches presented in this chapter.

S. Heinz, *Mathematical Modeling*, DOI 10.1007/978-3-642-20311-4_9,

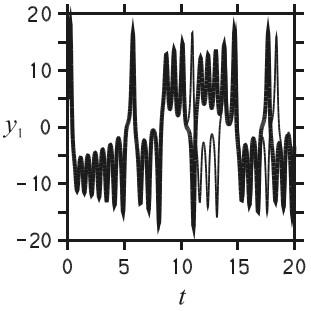



**Fig. 9.1.** A solution $y_1(t)$ of the Lorenz equations (9.1) combined with the model parameter $R = 28$. The *thick line* and the *thin line* present solutions for the initial values $(y_{10}, y_{20}, y_{30}) = (5, 5, 5)$ and $(y_{10}, y_{20}, y_{30}) = (5.01, 5, 5)$, respectively.

## 9.1 Motivation

**Weather Forecasting.** Weather forecasting is crucially relevant. Weather warnings are used to protect life and property. Temperature and precipitation forecasts are highly relevant to agriculture. Regarding everyday life, weather forecasts are used to find out what to wear on a particular day. Weather forecasting has to be performed of the basis of numerical solutions of complicated equations (systems of nonlinear coupled partial differential equations) that involve several variables (as the three velocity components in space and temperature). To see basic features of such systems of partial differential equations (like the weather predictability), it is helpful to consider highly simplified approximations to these equation systems – as given by the Lorenz (1963) model. The latter model is given by the equations

$$\frac{dy_1}{dt} = 10(y_2 - y_1), \tag{9.1a}$$

$$\frac{dy_2}{dt} = y_1(R - y_3) - y_2, \tag{9.1b}$$

$$\frac{dy_3}{dt} = y_1 y_2 - \frac{8}{3} y_3. \tag{9.1c}$$

Here, $y_1$ measures the strength and direction of atmospheric circulation, and $y_2$ and $y_2$ measure the horizontal and vertical temperature variation, respectively. The variable that essentially controls the dynamics of this equation system is $R$, which is proportional to the vertical temperature difference. Two solutions to these equations, which differ by a minor difference of the initial value for $y_1$, are shown in Fig. 9.1 (details about the Lorenz equations (9.1) and their numerical solutions can be found in Sect. 9.5). This figure illustrates that solutions of the Lorenz equations reveal a complicated behavior. Also, small variations of initial conditions may result in completely different solutions – which indicates that long-range weather forecasting may be impossible.

**Questions Considered.** The analysis of solutions of the Lorenz equations (9.1) leads to questions like:
- How can we determine different (chaotic and nonchaotic) solution regimes?
- How can we analyze the influence of variations of initial conditions?
- How can we characterize the asymptotic behavior of solutions?

The Lorenz equations do only represent one example for many problems that have to be described by (linear and nonlinear) systems of coupled equations for several variables. Thus, from a more general point of view there are questions like:
- How can we formulate laws for the multivariate evolution of several variables?
- Do all multivariate evolution equations have (convergent numerical) solutions?
- How can we analytically study multivariate evolution equations?

The latter and other questions will be addressed in this chapter.

## 9.2 Systems of First-Order Differential Equations

Techniques for the solution of linear systems and the analysis of basic features of solutions to nonlinear systems of first-order ordinary differential equations will be described in this section. This discussion will provide an appropriate basis for the developments to be performed in the following sections of this chapter.

### *9.2.1 Linear Systems of First-Order Differential Equations*

**Equations Considered.** The analysis of linear systems of ordinary differential equations is helpful because of two reasons: many problems (like the linearized pendulum equation: see Sect. 9.4.2) can be solved by linear equations, and linear equation systems can be used to analyze the solution features on nonlinear equation systems (like the Lotka-Volterra equations and Lorenz equations: see Sects. 9.3.3 and 9.5.2, respectively). Thus, let us consider the following linear equation system,

$$\frac{dy_1}{dt} = a_{11}y_1 + a_{12}y_2\,, \tag{9.2a}$$

$$\frac{dy_2}{dt} = a_{21}y_1 + a_{22}y_2\,. \tag{9.2b}$$

Here, $a_{11}$, $a_{12}$, $a_{21}$, and $a_{22}$ are constants. The solution of the equation system (9.2) requires initial values $y_1(0) = y_{10}$ and $y_2(0) = y_{20}$, where $y_{10}$ and $y_{20}$ are considered to be given parameters.

**Relation to Second-Order Equations.** A good way to find the solutions $y_1(t)$ and $y_2(t)$ to the equation system (9.2) is to exploit the relationship between this equation system and second-order differential equations. This relationship can be derived in the following way,

$$\frac{d^2y_1}{dt^2} = a_{11}\frac{dy_1}{dt} + a_{12}\frac{dy_2}{dt} = a_{11}\frac{dy_1}{dt} + a_{12}(a_{21}y_1 + a_{22}y_2)$$
$$= a_{11}\frac{dy_1}{dt} + a_{12}a_{21}y_1 + a_{22}\left(\frac{dy_1}{dt} - a_{11}y_1\right) = (a_{11} + a_{22})\frac{dy_1}{dt} - (a_{11}a_{22} - a_{12}a_{21})\,y_1. \tag{9.3}$$

In the first line, Eq. (9.2b) was applied to replace $dy_2 / dt$. In the second line, we used Eq. (9.2a) to replace $a_{12}\,y_2$. In the same way we find for $d^2y_2 / dt^2$

$$\frac{d^2y_2}{dt^2} = a_{21}\frac{dy_1}{dt} + a_{22}\frac{dy_2}{dt} = a_{21}(a_{11}y_1 + a_{12}y_2) + a_{22}\frac{dy_2}{dt}$$
$$= a_{22}\frac{dy_2}{dt} + a_{12}a_{21}y_2 + a_{11}\left(\frac{dy_2}{dt} - a_{22}y_2\right) = (a_{11} + a_{22})\frac{dy_2}{dt} - (a_{11}a_{22} - a_{12}a_{21})\,y_2. \tag{9.4}$$

Equations (9.3) and (9.4) can also be written

$$\frac{d^2y_1}{dt^2} + b\frac{dy_1}{dt} + c\,y_1 = 0, \tag{9.5a}$$

$$\frac{d^2y_2}{dt^2} + b\frac{dy_2}{dt} + c\,y_2 = 0, \tag{9.5b}$$

where the following abbreviations are applied,

$$b = -(a_{11} + a_{22}), \qquad c = a_{11}\,a_{22} - a_{12}\,a_{21}. \tag{9.6}$$

Equations (9.5a) and (9.5b) represent the same equation. Different solutions $y_1(t)$ and $y_2(t)$ of these equations are obtained by applying the initial values $y_1(0) = y_{10}$ and $y_2(0) = y_{20}$, and the initial derivatives $dy_1 / dt(0) = y'_{10}$ and $dy_2 / dt(0) = y'_{20}$ that are provided through Eq. (9.2).

$$y'_{10} = a_{11}y_{10} + a_{12}y_{20}\,, \tag{9.7a}$$

$$y'_{20} = a_{21}y_{10} + a_{22}y_{20}\,. \tag{9.7b}$$

Equations (9.5) correspond to the second-order equation (7.45). Accordingly, the solution of Eq. (7.45) derived in Chap. 7 can be used for the solution of the equation system (9.5), as will be shown in the next paragraph. Before doing this we will show that the relationship between a linear system of first-order equations and a linear second-order differential equation also can be used to write a linear second-order equation in terms of a system of first-order differential equations.

The latter can be seen by considering Eq. (7.45),

$$a\frac{d^2y}{dt^2}+b\frac{dy}{dt}+c\,y=0. \tag{9.8}$$

We set $y_1 = y$ and $y_2 = dy/dt$. Differentiation of $y_1$ and $y_2$ then provides

$$\frac{dy_1}{dt}=\frac{dy}{dt}=y_2, \tag{9.9a}$$

$$\frac{dy_2}{dt}=\frac{d^2y}{dt^2}=-\frac{b}{a}\frac{dy}{dt}-\frac{c}{a}y=-\frac{c}{a}y_1-\frac{b}{a}y_2, \tag{9.9b}$$

where Eq. (9.8) was applied. Hence, the second-order differential equation (9.8) can be represented as the system (9.9) of first-order equations. The initial values required for the solution of Eq. (9.9) are given by the initial values $y_1(0) = y(0)$ and $y_2(0) = dy/dt(0)$ that complete the second-order equation (9.8).

**Solution of the Equation System. I.** The solutions to Eqs. (9.2) can be derived via the solution (7.57) of the corresponding second-order equations (9.5). For $a = 1$ we obtain according to Eq. (7.57)

$$y_1=\frac{y'_{10}-r_2\,y_{10}}{r_1-r_2}e^{r_1t}-\frac{y'_{10}-r_1\,y_{10}}{r_1-r_2}e^{r_2t}, \tag{9.10a}$$

$$y_2=\frac{y'_{20}-r_2\,y_{20}}{r_1-r_2}e^{r_1t}-\frac{y'_{20}-r_1\,y_{20}}{r_1-r_2}e^{r_2t}. \tag{9.10b}$$

The initial derivatives $y'_{10}$ and $y'_{20}$ are given through Eqs. (9.7). The eigenvalues $r_1$ and $r_2$ are determined by the characteristic equation

$$0=r^2-(a_{11}+a_{22})r+a_{11}a_{22}-a_{12}\,a_{21}=(a_{11}-r)(a_{22}-r)-a_{12}\,a_{21}, \tag{9.11}$$

which follows from the use of $b = -(a_{11} + a_{22})$ and $c = a_{11}\,a_{22} - a_{12}\,a_{21}$ in the characteristic equation $r^2 + b\,r + c = 0$. Hence, the eigenvalues $r_1$ and $r_2$ are

$$r_1=r_S+r_D, \qquad r_2=r_S-r_D, \tag{9.12}$$

where $r_S$ and $r_D$ are given by

$$r_S=\frac{a_{11}+a_{22}}{2}, \tag{9.13a}$$

$$r_D=\frac{1}{2}\sqrt{(a_{11}+a_{22})^2-4(a_{11}a_{22}-a_{12}a_{21})}=\frac{1}{2}\sqrt{(a_{11}-a_{22})^2+4a_{12}a_{21}}. \tag{9.13b}$$

**Solution of the Equation System. II.** Let us directly solve the equation system (9.2) to check the validity of the solutions (9.10). Such a solution can be found efficiently by making use of vector and matrix notation. We write Eq. (9.2) as

$$\frac{d\boldsymbol{y}}{dt}=\boldsymbol{A}\,\boldsymbol{y}. \tag{9.14}$$

Here, the vector $\boldsymbol{y}$ and the matrix $\boldsymbol{A}$ are given by

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \qquad \boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}. \tag{9.15}$$

To solve Eq. (9.14) we assume according to Eq. (9.10) an exponential solution,

$$\boldsymbol{y} = \boldsymbol{c}\, e^{rt}, \tag{9.16}$$

where the constant vector **c** is given by

$$\boldsymbol{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}. \tag{9.17}$$

The use of Eq. (9.16) in Eq. (9.14) then results in

$$r\,\boldsymbol{c}\, e^{rt} = \boldsymbol{A}\,\boldsymbol{c}\, e^{rt}. \tag{9.18}$$

Upon cancelling the nonzero exponential function we obtain

$$(\boldsymbol{A} - r\,\mathbf{I})\,\boldsymbol{c} = 0. \tag{9.19}$$

Here, **I** is the 2×2 identity matrix,

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{9.20}$$

The identity matrix **I** has the property $\mathbf{I}\,\boldsymbol{c} = \boldsymbol{c}$. The inverse matrix of $\boldsymbol{A} - r\,\mathbf{I}$ will exist if the determinant $\det(\boldsymbol{A} - r\,\mathbf{I})$ is nonzero. In this case, we do only obtain trivial solutions $\boldsymbol{c} = \mathbf{0}$ according to Eq. (9.19). Thus, the condition to obtain non-trivial solutions $\boldsymbol{c}$ is given by $\det(\boldsymbol{A} - r\,\mathbf{I}) = 0$, i.e.,

$$\det(\boldsymbol{A} - r\,\mathbf{I}) = \begin{vmatrix} a_{11} - r & a_{12} \\ a_{21} & a_{22} - r \end{vmatrix} = 0. \tag{9.21}$$

The latter constraint can be also written

$$0 = (a_{11} - r)(a_{22} - r) - a_{12}\,a_{21} = r^2 - (a_{11} + a_{22})\,r + a_{11}a_{22} - a_{12}\,a_{21}. \tag{9.22}$$

The solution of this quadratic equation for $r$ reveals that the two eigenvalues $r_1$ and $r_2$ obtained in this way agree with the eigenvalues $r_1$ and $r_2$ given by Eq. (9.12). To specify the solution (9.16) we have to use the eigenvalues $r_1$ and $r_2$ in Eq. (9.19) to determine the corresponding eigenvectors $\boldsymbol{c}^{(1)}$ and $\boldsymbol{c}^{(2)}$. This constraint provides the equations

$$\begin{pmatrix} a_{11} - r_1 & a_{12} \\ a_{21} & a_{22} - r_1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}^{(1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{9.23a}$$

$$\begin{pmatrix} a_{11} - r_2 & a_{12} \\ a_{21} & a_{22} - r_2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}^{(2)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{9.23b}$$

The conditions for the first eigenvector $\boldsymbol{c}^{(1)}$ can be written

$$0 = (a_{11} - r_1)c_1^{(1)} + a_{12}\,c_2^{(1)}, \tag{9.24a}$$

$$0 = a_{21}\,c_1^{(1)} + (a_{22} - r_1)c_2^{(1)}. \tag{9.24b}$$

Equation (9.24a) can be used to express $c_2^{(1)}$ in terms of $c_1^{(1)}$,

$$c_2^{(1)} = -\frac{a_{11} - r_1}{a_{12}}c_1^{(1)}. \tag{9.25}$$

The use of this expression in Eq. (9.24b) provides a relation for $c_1^{(1)}$,

$$0 = \left[a_{21} - \frac{1}{a_{12}}(a_{11} - r_1)(a_{22} - r_1)\right]c_1^{(1)} = \left[(a_{11} - r_1)(a_{22} - r_1) - a_{12}\,a_{21}\right]c_1^{(1)}. \tag{9.26}$$

The eigenvector $r_1$ solves the characteristic Eq (9.22). Thus, the bracket term in the latter relation is equal to zero, which means that there is no constraint on $c_1^{(1)}$. Hence, we may assume that $c_1^{(1)} = c_1$, where $c_1$ is an open parameter. The eigenvector $\boldsymbol{c}^{(1)}$ can be written then $\boldsymbol{c}^{(1)} = c_1\,\boldsymbol{u}_1$, where

$$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ -\dfrac{a_{11} - r_1}{a_{12}} \end{pmatrix}. \tag{9.27}$$

The corresponding condition for the second eigenvectors $\boldsymbol{c}^{(2)}$ can be derived in the same way. According to Eq. (9.23b) we have

$$0 = (a_{11} - r_2)c_1^{(2)} + a_{12}\,c_2^{(2)}, \tag{9.28a}$$

$$0 = a_{21}\,c_1^{(2)} + (a_{22} - r_2)c_2^{(2)}. \tag{9.28b}$$

Using the relation for $c_1^{(2)}$ implied by Eq. (9.28a) in the second relation leads to

$$0 = \left[-a_{21}\frac{a_{12}}{a_{11} - r_2} + a_{22} - r_2\right]c_2^{(2)} = \left[(a_{11} - r_2)(a_{22} - r_2) - a_{12}\,a_{21}\right]\,c_2^{(2)}. \tag{9.29}$$

This relation does not imply a condition on $c_2^{(2)}$ because the eigenvalue $r_2$ satisfies the characteristic equation (9.22). By setting $c_2^{(2)} = c_2$, where $c_2$ is an open parameter, we find by means of Eq. (9.28a)

$$c_1^{(2)} = -\frac{a_{12}}{a_{11} - r_2}c_2. \tag{9.30}$$

Hence, $\boldsymbol{c}^{(2)}$ can be written $\boldsymbol{c}^{(2)} = c_2\,\boldsymbol{u}_2$, where

$$\boldsymbol{u}_2 = \begin{pmatrix} -\dfrac{a_{12}}{a_{11} - r_2} \\ 1 \end{pmatrix}. \tag{9.31}$$

By accounting for the two possible exponential solutions we can write the solution of the equation system (9.14) as

$$\boldsymbol{y} = c_1 \, \boldsymbol{u}_1 \, e^{r_1 t} + c_2 \, \boldsymbol{u}_2 \, e^{r_2 t}, \tag{9.32}$$

which means according to Eqs. (9.27) and (9.31) that

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ -\dfrac{a_{11} - r_1}{a_{12}} \end{pmatrix} e^{r_1 t} + c_2 \begin{pmatrix} -\dfrac{a_{12}}{a_{11} - r_2} \\ 1 \end{pmatrix} e^{r_2 t}. \tag{9.33}$$

This solution has to satisfy the initial conditions, which implies two equations for $c_1$ and $c_2$,

$$y_{10} = c_1 - \frac{a_{12}}{a_{11} - r_2} c_2, \qquad y_{20} = -\frac{a_{11} - r_1}{a_{12}} c_1 + c_2. \tag{9.34}$$

The use of the second relation for $c_2$ in the first condition, and the use of the first relation for $c_1$ in the second relation implies

$$y_{10} = c_1 - \frac{a_{12}}{a_{11} - r_2} \left[ y_{20} + \frac{a_{11} - r_1}{a_{12}} c_1 \right] = c_1 \left( 1 - \frac{a_{11} - r_1}{a_{11} - r_2} \right) - \frac{a_{12}}{a_{11} - r_2} y_{20}, \tag{9.35a}$$

$$y_{20} = -\frac{a_{11} - r_1}{a_{12}} \left[ y_{10} + \frac{a_{12}}{a_{11} - r_2} c_2 \right] + c_2 = c_2 \left( 1 - \frac{a_{11} - r_1}{a_{11} - r_2} \right) - \frac{a_{11} - r_1}{a_{12}} y_{10}. \tag{9.35b}$$

Correspondingly, $c_1$ and $c_2$ are given by

$$c_1 = \frac{y_{10}(a_{11} - r_2) + a_{12} y_{20}}{r_1 - r_2} = \frac{y'_{10} - r_2 \, y_{10}}{r_1 - r_2}, \tag{9.36a}$$

$$c_2 = \frac{a_{11} - r_2}{a_{12}(r_1 - r_2)} \left[ a_{12} \, y_{20} + (a_{11} - r_1) \, y_{10} \right] = \frac{a_{11} - r_2}{a_{12}} \frac{y'_{10} - r_1 \, y_{10}}{r_1 - r_2}, \tag{9.36b}$$

where the definitions (9.7) of initial derivatives are used as abbreviations. With these expressions we can write the solution of Eq. (9.14)

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{y'_{10} - r_2 \, y_{10}}{r_1 - r_2} \begin{pmatrix} 1 \\ -\dfrac{a_{11} - r_1}{a_{12}} \end{pmatrix} e^{r_1 t} - \frac{y'_{10} - r_1 \, y_{10}}{r_1 - r_2} \begin{pmatrix} 1 \\ -\dfrac{a_{11} - r_2}{a_{12}} \end{pmatrix} e^{r_2 t}. \tag{9.37}$$

**Consistency of Solutions.** This solution provides $y_1$ as given by Eq. (9.10a). To show that $y_2$ given by Eq. (9.37) agrees with Eq. (9.10b) we do the following: The initial value $y_{20}$ and initial derivative $y'_{20}$ implied by Eq. (9.37) are given by

$$y_{20} = \frac{P - Q}{r_1 - r_2}, \qquad y'_{20} = \frac{r_1 P - r_2 Q}{r_1 - r_2}. \tag{9.38}$$

**Table 9.1** Cases considered for the illustration of solutions of the linear equation system (9.42).

| | $a_{11}$ | $a_{12}$ | $a_{21}$ | $r_1$ | $r_2$ |
|---|---|---|---|---|---|
| Case 1: Real unequal eigenvalues of the same sign | –1 | 0.5 | 0.5 | –0.5 | –1.5 |
| Case 2: Real unequal eigenvalues of opposite sign | –1 | 2 | 2 | 1 | –3 |
| Case 3: Real equal eigenvalues | –1 | 0 | 0 | –1 | –1 |
| Case 4: Complex eigenvalues | –1 | 1 | –1 | –1+i | –1–i |
| Case 5: Pure imaginary eigenvalues | 0 | 1 | –1 | i | –i |

Here, we used the abbreviations

$$P = -\frac{a_{11} - r_1}{a_{12}}(y'_{10} - r_2\, y_{10}), \qquad Q = -\frac{a_{11} - r_2}{a_{12}}(y'_{10} - r_1\, y_{10}). \tag{9.39}$$

The use of $P = (r_1 - r_2)\, y_{20} + Q$ and $Q = P - (r_1 - r_2)\, y_{20}$ according to Eq. (9.39) enables us to write

$$y'_{20} = \frac{r_1\left[Q + (r_1 - r_2)\, y_{20}\right] - r_2\, Q}{r_1 - r_2} = \frac{r_1\, P - r_2\left[P - (r_1 - r_2)\, y_{20}\right]}{r_1 - r_2}. \tag{9.40}$$

The latter two relations between $Q$ and $P$ with $y'_{20}$ can be used to obtain

$$y'_{20} - r_1\, y_{20} = Q = -\frac{a_{11} - r_2}{a_{12}}(y'_{10} - r_1\, y_{10}), \tag{9.41a}$$

$$y'_{20} - r_2\, y_{20} = P = -\frac{a_{11} - r_1}{a_{12}}(y'_{10} - r_2\, y_{10}). \tag{9.41b}$$

The combination of these relations with Eq. (9.37) recovers the solution (9.10b).

### *9.2.2 Features of Solutions of Linear First-Order Systems*

**Example.** Let us illustrate some characteristic features of solutions of the linear equation system (9.2). For simplicity we assume that $a_{11} = a_{22}$,

$$\frac{dy_1}{dt} = a_{11} y_1 + a_{12} y_2\,, \tag{9.42a}$$

$$\frac{dy_2}{dt} = a_{21} y_1 + a_{11} y_2\,. \tag{9.42b}$$

According to Eq. (9.12), the eigenvalues $r_1$ and $r_2$ are then given by the relations

$$r_1 = a_{11} + \sqrt{a_{12} a_{21}}\,, \qquad r_2 = a_{11} - \sqrt{a_{12} a_{21}}\,. \tag{9.43}$$

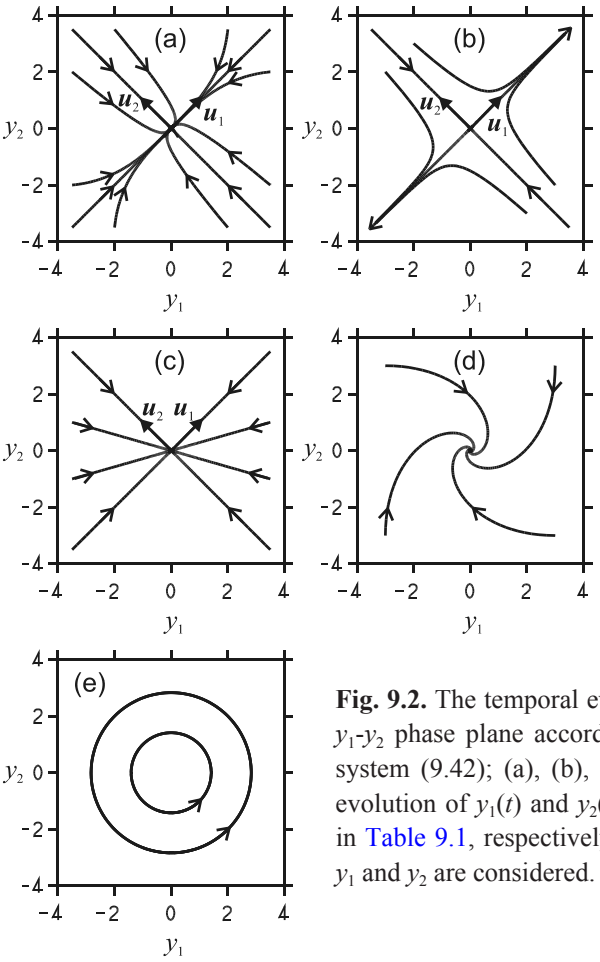



**Fig. 9.2.** The temporal evolution of $y_1$ and $y_2$ in the $y_1$-$y_2$ phase plane according to the linear equation system (9.42); (a), (b), (c), (d), and (e) show the evolution of $y_1(t)$ and $y_2(t)$ for the five cases given in Table 9.1, respectively. Several initial values of $y_1$ and $y_2$ are considered.

Five sets of model parameters $a_{11}$, $a_{12}$, and $a_{21}$ are specified in Table 9.1. The cases considered correspond to five characteristic types of the eigenvalues $r_1$ and $r_2$. The solutions of the equation system (9.42) were obtained numerically. The evolution of $y_1$ and $y_2$ in time is shown in Fig. 9.2 in the $y_1$-$y_2$ phase plane for several initial values $y_{10}$ and $y_{20}$. Such curves can be seen as the trajectory of a particle moving with a velocity $d\boldsymbol{y}/dt = \boldsymbol{A}\,\boldsymbol{y}$.

**Analytical Solutions.** The analytical solutions for the five cases considered are given by Eq. (9.32). To prepare the discussion of specific cases we will calculate the solution for the first three cases. Neither case 4 nor 5 will be involved here: these cases require rewritings of the solution (9.32) to have real-valued solutions (see Chap. 7). To cover the first three cases we set $a_{12} = a_{21} = \varepsilon$. Here, $\varepsilon$ represents a positive parameter. For the cases 1, 2, and 3 we have the values $\varepsilon = (0.5, 2, 0)$,

respectively. According to Eq. (9.43), the eigenvalues are then given by

$$r_1 = -1 + \varepsilon, \qquad r_2 = -1 - \varepsilon. \tag{9.44}$$

Equations (9.27) and (9.31) imply for the eigenvectors

$$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ -\dfrac{a_{11} - r_1}{a_{12}} \end{pmatrix} = \begin{pmatrix} 1 \\ -\dfrac{-\varepsilon}{\varepsilon} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \boldsymbol{u}_2 = \begin{pmatrix} -\dfrac{a_{12}}{a_{11} - r_2} \\ 1 \end{pmatrix} = \begin{pmatrix} -\dfrac{\varepsilon}{\varepsilon} \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \tag{9.45}$$

According to Eq. (9.36), the coefficients $c_1$ and $c_2$ are given by

$$c_1 = \frac{a_{11} y_{10} + a_{12} y_{20} - r_2\, y_{10}}{r_1 - r_2} = \frac{-y_{10} + \varepsilon\, y_{20} + (1+\varepsilon)\, y_{10}}{2\varepsilon} = \frac{y_{20} + y_{10}}{2}, \tag{9.46a}$$

$$c_2 = \frac{a_{11} - r_2}{a_{12}} \frac{a_{11} y_{10} + a_{12} y_{20} - r_1\, y_{10}}{r_1 - r_2} = \frac{\varepsilon}{\varepsilon} \frac{-y_{10} + \varepsilon\, y_{20} + (1-\varepsilon)\, y_{10}}{2\varepsilon} = \frac{y_{20} - y_{10}}{2}. \tag{9.46b}$$

Thus, the solution (9.32) reads

$$\boldsymbol{y} = c_1 \boldsymbol{u}_1 e^{r_1 t} + c_2 \boldsymbol{u}_2 e^{r_2 t} = \frac{y_{20} + y_{10}}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-(1-\varepsilon)t} + \frac{y_{20} - y_{10}}{2} \begin{pmatrix} -1 \\ 1 \end{pmatrix} e^{-(1+\varepsilon)t}. \tag{9.47}$$

In terms of $y_1$ and $y_2$, this solution can be written

$$y_1 = \frac{y_{20} + y_{10}}{2} e^{-(1-\varepsilon)t} - \frac{y_{20} - y_{10}}{2} e^{-(1+\varepsilon)t}, \tag{9.48a}$$

$$y_2 = \frac{y_{20} + y_{10}}{2} e^{-(1-\varepsilon)t} + \frac{y_{20} - y_{10}}{2} e^{-(1+\varepsilon)t}. \tag{9.48b}$$

To see the relation between $y_2$ on $y_1$, which determines the trajectory in the $y_1$-$y_2$ phase plane, we consider the sum and the difference of these two expressions,

$$y_2 + y_1 = (y_{20} + y_{10}) e^{-(1-\varepsilon)t}, \qquad y_2 - y_1 = (y_{20} - y_{10}) e^{-(1+\varepsilon)t}. \tag{9.49}$$

**Case 1.** The first case considers two real unequal eigenvalues of the same sign. For this case ($\varepsilon = 0.5$), the solution (9.47) reads

$$\boldsymbol{y} = c_1 \boldsymbol{u}_1 e^{-0.5t} + c_2 \boldsymbol{u}_2 e^{-1.5t} = e^{-0.5t} \left( c_1 \boldsymbol{u}_1 + c_2 \boldsymbol{u}_2 e^{-t} \right). \tag{9.50}$$

Examples for the evolution of $y_1$ and $y_2$ in time are shown in Fig. 9.2a for different initial values. All trajectories are attracted by the equilibrium solution (0, 0). For two real unequal eigenvalues that are positive one finds the opposite feature that all trajectories increase their distance to (0, 0). The trajectories are aligned with the eigenvectors. There is, however, a difference in the behavior of trajectories. The term $c\, \boldsymbol{u}_2\, e^{-t}$ in Eq. (9.50) is small compared to $c\, \boldsymbol{u}_1$ for sufficiently large $t$. Thus, the trajectories tend toward $\boldsymbol{u}_1$ before they reach the equilibrium point (0, 0).

The trajectories are characterized by Eq. (9.49), where $\varepsilon = 0.5$. For $t \to \infty$ we find $y_2 = \pm y_1$, which agrees with the eigenvectors (9.45).

**Case 2.** The second case considers two real unequal eigenvalues of opposite sign. The solution (9.47) reads for this case

$$\boldsymbol{y} = c_1 \boldsymbol{u}_1 e^{t} + c_2 \boldsymbol{u}_2 e^{-3t} = e^{t}\left(c_1 \boldsymbol{u}_1 + c_2 \boldsymbol{u}_2 e^{-4t}\right). \tag{9.51}$$

Examples for the evolution of $y_1$ and $y_2$ are shown in Fig. 9.2b for different initial values. The behavior of trajectories can be explained by considering the solution (9.51). As given for the first case, trajectories tend toward $\boldsymbol{u}_1$ because the $\boldsymbol{u}_2$ term in Eq. (9.51) becomes negligible compared to the $\boldsymbol{u}_1$ term for large $t$. The difference to the first case is given by the fact that the solution tends (due to the positive eigenvalue) to infinity after reaching $\boldsymbol{u}_1$. The relation between $y_1$ and $y_2$ is determined by Eq. (9.49), where $\varepsilon = 2$. For $t \to \infty$ this relation provides $y_2 = \pm y_1$ in agreement with the eigenvectors (9.45).

**Case 3.** The third case considers two real equal eigenvalues. The solution for this case can be found by considering the limit $\varepsilon \to 0$ of Eq. (9.47),

$$\boldsymbol{y} = \left[\frac{y_{20} + y_{10}}{2}\begin{pmatrix}1\\1\end{pmatrix} + \frac{y_{20} - y_{10}}{2}\begin{pmatrix}-1\\1\end{pmatrix}\right] e^{-t} = \begin{pmatrix}y_{10}\\y_{20}\end{pmatrix} e^{-t}. \tag{9.52}$$

Examples for the evolution of $y_1$ and $y_2$ in time are shown in Fig. 9.2c. Expression (9.52) explains the difference to the cases 1 and 2: trajectories do not tend toward $\boldsymbol{u}_1$ because the bracket term is independent of time. The path of trajectories can be derived from Eq. (9.52),

$$y_2 = \frac{y_{20}}{y_{10}} y_1 . \tag{9.53}$$

Hence, every trajectory lies on a straight line through the origin.

**Case 4.** The fourth case considers two complex eigenvalues. The eigenvectors are also complex. Figure 9.2d shows that all trajectories tend toward the equilibrium point (0, 0). The simplest way to see in which way the equilibrium is established is to analyze the consequences of the equation system (9.42) directly,

$$\frac{dy_1}{dt} = -y_1 + y_2 , \tag{9.54a}$$

$$\frac{dy_2}{dt} = -y_1 - y_2 . \tag{9.54b}$$

We multiply Eq. (9.54a) by $2\,y_1$ and Eq. (9.54b) by $2\,y_2$, and we take the sum of both equations,

$$\frac{d}{dt}(y_1^{\,2} + y_2^{\,2}) = -2(y_1^{\,2} + y_2^{\,2}). \tag{9.55}$$

This relation was obtained by making use of the identity $dy_1^{\,2}/dt = 2\,y_1\,dy_1/dt$, and

a corresponding relation for $y_2$. The solution of this equation is given by

$$y_1^2 + y_2^2 = (y_{10}^2 + y_{20}^2)e^{-2t}. \tag{9.56}$$

Therefore, the trajectories represent circles with a radius that decreases in time, this means the trajectories are spirals.

**Case 5.** The fifth case considers pure imaginary eigenvalues. Figure 9.2e shows that the trajectories are given by circles in this case. Evidence for these trajectories can be obtained by the consideration of the equation system (9.42) for this case,

$$\frac{dy_1}{dt} = y_2, \tag{9.57a}$$

$$\frac{dy_2}{dt} = -y_1. \tag{9.57b}$$

As for the fourth case, we multiply Eq. (9.57a) by $2y_1$, Eq. (9.57b) by $2y_2$, and we take the sum of both equations,

$$\frac{d}{dt}(y_1^2 + y_2^2) = 0. \tag{9.58}$$

Hence, the trajectories are indeed circles,

$$y_1^2 + y_2^2 = y_{10}^2 + y_{20}^2. \tag{9.59}$$

The critical point (0, 0) is called a center.

**Summary.** The features of solutions of systems of first-order linear differential equations can be summarized in the following way (Boyce and DiPrima 2009). There are three possibilities for the evolution of trajectories:

a) Trajectories approach the equilibrium point as $t \to \infty$. This behavior is seen if the eigenvalues are real and negative or complex with real negative part. Such a system is called asymptotically stable.
b) Trajectories remain bounded but they do not approach the origin. This behavior appears if the eigenvalues are pure imaginary. Such a system is called stable.
c) Trajectories become unbounded as $t \to \infty$. Such a behavior is seen if at least one eigenvalue is positive or if the eigenvalues have a positive real part. Such a system is called unstable.

### *9.2.3 Analysis of Nonlinear Equation Systems*

**Nonlinear Equation System.** The analysis of nonlinear equation systems is more difficult than the analysis of linear systems because nonlinear systems can hardly be solved analytically. To illustrate the way of analyzing the behavior of

nonlinear systems, let us consider an equation system that will be used also for the discussion of population ecology processes, see Sect. 9.3,

$$\frac{dy_1}{dt} = y_1\left(a_1 + b_1 y_1 + c_1 y_2\right), \tag{9.60a}$$

$$\frac{dy_2}{dt} = y_2\left(a_2 + b_2 y_2 + c_2 y_1\right). \tag{9.60b}$$

Here, $a_1$, $b_1$, $c_1$ and $a_2$, $b_2$, $c_2$ are any positive or negative constants. One approach to analyze the nonlinear equation system (9.60) will be discussed in the following: The idea is to determine the equilibrium points and to analyze the solution sufficiently close to the equilibrium points such that the nonlinear equation system can be approximated by a linear equation system that can be solved.

**Equilibrium Points.** An equilibrium solution ($Y_1$, $Y_2$) of Eqs. (9.60) is defined by $Y_1$ and $Y_2$ values so that $dy_1 / dt = dy_2 / dt = 0$. Therefore, equilibrium solutions are defined by the conditions

$$0 = y_1\left(a_1 + b_1 y_1 + c_1 y_2\right), \tag{9.61a}$$

$$0 = y_2\left(a_2 + b_2 y_2 + c_2 y_1\right). \tag{9.61b}$$

The equation system (9.60) provides four equilibrium points, which are given by

$$\begin{aligned} &(Y_1, Y_2) = (0,0), \qquad (Y_1, Y_2) = \left(0, -\frac{a_2}{b_2}\right), \qquad (Y_1, Y_2) = \left(-\frac{a_1}{b_1}, 0\right), \\ &(Y_1, Y_2) = \left(\frac{a_2 c_1 - a_1 b_2}{b_1 b_2 - c_1 c_2}, \frac{a_1 c_2 - a_2 b_1}{b_1 b_2 - c_1 c_2}\right). \end{aligned} \tag{9.62}$$

The validity of the first three equilibrium points can be easily seen. The last equilibrium point ensures that both parenthesis terms are equal to zero. The validity of this claim may be proven by using this point in the parenthesis terms of Eqs. (9.61a) and (9.61b), respectively,

$$\begin{aligned} a_1 + b_1 y_1 + c_1 y_2 &= a_1 + b_1 \frac{a_2 c_1 - a_1 b_2}{b_1 b_2 - c_1 c_2} + c_1 \frac{a_1 c_2 - a_2 b_1}{b_1 b_2 - c_1 c_2} \\ &= a_1 (b_1 b_2 - c_1 c_2) + b_1 (a_2 c_1 - a_1 b_2) + c_1 (a_1 c_2 - a_2 b_1) = 0, \end{aligned} \tag{9.63a}$$

$$\begin{aligned} a_2 + b_2 y_2 + c_2 y_1 &= a_2 + b_2 \frac{a_1 c_2 - a_2 b_1}{b_1 b_2 - c_1 c_2} + c_2 \frac{a_2 c_1 - a_1 b_2}{b_1 b_2 - c_1 c_2} \\ &= a_2 (b_1 b_2 - c_1 c_2) + b_2 (a_1 c_2 - a_2 b_1) + c_2 (a_2 c_1 - a_1 b_2) = 0. \end{aligned} \tag{9.63b}$$

**Near-Equilibrium Equation System.** The next step is to study the behavior of small deviations of the solution from the equilibrium points. Small deviations from the equilibrium points are defined by setting

$$y_1 = Y_1 + v_1, \qquad y_2 = Y_2 + v_2. \tag{9.64}$$

Here, $(Y_1, Y_2)$ represent the coordinates of any equilibrium point, and the functions $(v_1, v_2)$ are small deviations from this equilibrium point. By replacing $y_1$ and $y_2$ in Eq. (9.60) by the latter expressions we obtain

$$\begin{aligned}\frac{dv_1}{dt} &= (Y_1+v_1)\left[a_1+b_1(Y_1+v_1)+c_1(Y_2+v_2)\right] = Y_1\left(a_1+b_1Y_1+c_1Y_2\right)\\ &\quad + v_1\left[a_1+b_1Y_1+c_1Y_2+b_1Y_1\right]+v_2\,c_1Y_1+v_1(b_1v_1+c_1v_2),\end{aligned} \tag{9.65a}$$

$$\begin{aligned}\frac{dv_2}{dt} &= (Y_2+v_2)\left[a_2+b_2(Y_2+v_2)+c_2(Y_1+v_1)\right] = Y_2\left(a_2+b_2Y_2+c_2Y_1\right)\\ &\quad + v_1\,c_2Y_2+v_2\left[a_2+b_2Y_2+c_2Y_1+b_2Y_2\right]+v_2(b_2v_2+c_2v_1).\end{aligned} \tag{9.65b}$$

The first terms on the right-hand sides of these relations are zero because $Y_1$ and $Y_2$ are equilibrium solutions. The last terms are quadratic in $v_1$ and $v_2$. The latter terms can be neglected because $v_1$ and $v_2$ are assumed to be small. In this way, we obtain a linear equation system for $v_1$ and $v_2$,

$$\frac{dv_1}{dt} = v_1\left[a_1+b_1Y_1+c_1Y_2+b_1Y_1\right]+v_2\,c_1Y_1, \tag{9.66a}$$

$$\frac{dv_2}{dt} = v_1\,c_2Y_2+v_2\left[a_2+b_2Y_2+c_2Y_1+b_2Y_2\right]. \tag{9.66b}$$

This linear equation system can be solved in terms of the solutions provided in Sect. 9.2.1.

**Generalization.** The linearization of the nonlinear equation system described in the preceding paragraph can be applied to any nonlinear equation system. The latter fact can be demonstrated by considering the nonlinear system

$$\frac{dy_1}{dt} = F_1(y_1, y_2), \tag{9.67a}$$

$$\frac{dy_2}{dt} = F_2(y_1, y_2), \tag{9.67b}$$

where $F_1$ and $F_2$ can be any functions of $y_1$ and $y_2$. The Taylor expansion of $F_1$ and $F_2$ at an equilibrium point $(Y_1, Y_2)$ is then given by

$$\frac{dy_1}{dt} = F_1(Y_1,Y_2)+\frac{\partial F_1}{\partial y_1}(Y_1,Y_2)\left(y_1-Y_1\right)+\frac{\partial F_1}{\partial y_2}(Y_1,Y_2)\left(y_2-Y_2\right), \tag{9.68a}$$

$$\frac{dy_2}{dt} = F_2(Y_1,Y_2)+\frac{\partial F_2}{\partial y_1}(Y_1,Y_2)\left(y_1-Y_1\right)+\frac{\partial F_2}{\partial y_2}(Y_1,Y_2)\left(y_2-Y_2\right), \tag{9.68b}$$

where nonlinear powers of $y_1 - Y_1$ and $y_2 - Y_2$ are neglected (which is justified for sufficiently small deviations from the equilibrium point). The functions $F_1$ and $F_2$ are equal to zero at the equilibrium points, i.e., we have $F_1(Y_1, Y_2) = F_2(Y_1, Y_2) = 0$.

Hence, we obtain a linear equation system in $y_1 - Y_1$ and $y_2 - Y_2$,

$$\frac{d}{dt}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial F_1}{\partial y_1}(Y_1, Y_2) & \frac{\partial F_1}{\partial y_2}(Y_1, Y_2) \\ \frac{\partial F_2}{\partial y_1}(Y_1, Y_2) & \frac{\partial F_2}{\partial y_2}(Y_1, Y_2) \end{pmatrix} \begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix}. \tag{9.69}$$

By calculating the partial derivatives involved, this equation system can be used to recover the equation system (9.66), where $v_1 = y_1 - Y_1$ and $v_2 = y_2 - Y_2$.

**Application.** Examples for the application of this approach will be discussed in Sect. 9.3. These examples illustrate the benefits of such linear stability analyses. However, such analyses are not always successful. The latter is the case if the linear system that characterizes the system behavior in the neighborhood of an equilibrium point has two pure imaginary eigenvalues such that the trajectories are closed curves (ellipses). In that case, small disturbances given by nonlinear terms generate positive or negative real parts of the complex eigenvalues. Depending on the sign of these real parts, the nonlinear system may be asymptotially stable or unstable. Therefore, the analysis of the corresponding linear system does not allow in this case to decide whether or not the nonlinear system is asymptotially stable. A procedure for handling this problem will be discussed in Sect. 9.4 where Liapunov's second method is explained.

## 9.3 Population Ecology: Species Interactions

As a first application of the mathematical concepts presented in Sect. 9.2, let us consider the modeling of the multivariate evolution of several populations. This problem will be addressed such that the concepts presented for a single population in Chap. 7 are extended by the consideration of the interaction of several species. The discussions in Chap. 7 showed that there is no unique law of population ecology, but (depending on the definition of the population density function) there are many possibilities for formulating equations for population dynamics. In the following, we will consider modeling approaches that extend the logistic growth model for a single population.

### *9.3.1 Multivariate Population Dynamics Equations*

**Multivariate Evolution.** The following discussion of some basic features of the multivariate evolution of populations will be based on the nonlinear equation

system (9.60), which was analyzed mathematically in Sect. 9.2.3,

$$\frac{dy_1}{dt} = y_1\left(a_1 + b_1 y_1 + c_1 y_2\right), \tag{9.70a}$$

$$\frac{dy_2}{dt} = y_2\left(a_2 + b_2 y_2 + c_2 y_1\right). \tag{9.70b}$$

The structure of this model enables the consideration of a variety of types of species interactions. We will discuss two examples in the following.

**Competition for Food.** As a first example, let us consider the competition for food by two species that do not prey on each other. An example is given by two species of fish (bluegill and redear) in a pond. The equations considered for this case are given by

$$\frac{dy_1}{dt} = y_1\left(|\, a_1 \,| - |\, b_1 \,|\, y_1 - |\, c_1 \,|\, y_2\right), \tag{9.71a}$$

$$\frac{dy_2}{dt} = y_2\left(|\, a_2 \,| - |\, b_2 \,|\, y_2 - |\, c_2 \,|\, y_1\right). \tag{9.71b}$$

For the case that there is no interaction between species, i.e., $c_1 = c_2 = 0$, these equations represent logistic growth models for $y_1$ and $y_2$. The interaction terms that involve $c_1$ and $c_2$ do appear here with negative coefficients. In this way, we model the food reduction for one species due to the food consumption of the other species. This model will be analyzed in Sect. 9.3.2.

**Predator-Prey Interactions.** As a second example, we consider predator-prey interactions (e.g., foxes and rabbits in a closed forest). We assume that $y_1$ refers to the prey, and $y_2$ refers to the predator. The equations for this case are given by

$$\frac{dy_1}{dt} = y_1\left(|\, a_1 \,| - |\, b_1 \,|\, y_1 - |\, c_1 \,|\, y_2\right), \tag{9.72a}$$

$$\frac{dy_2}{dt} = y_2\left(-|\, a_2 \,| + |\, c_2 \,|\, y_1\right). \tag{9.72b}$$

The prey equation (9.72a) has the same structure as Eq. (9.71a): we have a logistic model with an interaction term that is proportional to $c_1$. A nonzero $c_1$ accounts for the reduction of prey due to predators. On the other hand, the predator equation (9.72b) differs from Eq. (9.71b). The predator will die out in the absence of the prey. The consideration of a self-limiting factor (i.e., a nonzero $b_2$) does not make sense in this scenario. For a nonzero coefficient $c_2$, the positive last term describes the increase of the predator population due to the consumption of prey. Equations (9.72) represent the famous Lotka-Volterra equations, which are extended here by the consideration of the self-limiting contribution related to the use of a nonzero $b_1$. This model will be analyzed in Sect. 9.3.3.

### 9.3.2 Competition for Food

**Model Considered.** For an analysis of the equation system (9.71) it is helpful to consider more specific equations. The parameters $b_1$ and $b_2$ normalize the other model parameters: see Eq. (9.70). Thus, we may set $b_1 = b_2 = -1$ (the negative sign is considered according to Eq. (9.71)). The equilibrium values (9.62) show that the second and third equilibrium values of $Y_1$ and $Y_2$ are given then by $a_1$ and $a_2$, respectively. We will assume that $a_1 = a_2 = 1$, which corresponds to the consideration of normalized population values. Regarding $c_1$ and $c_2$ we assume that $c_1 = c_2 = -d$, where $d$ is a non-negative number. The model considered is then given by

$$\frac{dy_1}{dt} = y_1\left(1 - y_1 - d\, y_2\right), \tag{9.73a}$$

$$\frac{dy_2}{dt} = y_2\left(1 - y_2 - d\, y_1\right). \tag{9.73b}$$

The equations look similar, which leads to the question regarding the difference between them. To address this question we calculate the ratio between $y_2$ and $y_1$,

$$\begin{aligned}\frac{d}{dt}\frac{y_2}{y_1} &= \frac{1}{y_1}\frac{dy_2}{dt} - \frac{y_2}{{y_1}^2}\frac{dy_1}{dt} = \frac{y_2}{y_1}\left(1 - y_2 - d\, y_1\right) - \frac{y_2}{{y_1}^2}y_1\left(1 - y_1 - d\, y_2\right) \\ &= (1-d)\frac{y_2}{y_1}\left(y_1 - y_2\right) = (1-d)\left(1 - \frac{y_2}{y_1}\right)y_2 .\end{aligned} \tag{9.74}$$

This relation shows that $y_2 / y_1$ is constant (in particular, we have $y_2 / y_1 = y_{20} / y_{10}$) under two conditions: for $d = 1$ and for the case that $y_1$ and $y_2$ have the same initial condition. The conclusion for the latter case can be seen, for example, by writing Eq. (9.74) in a discrete formulation. Such a representation shows that there is never a change of $y_2 / y_1$. The latter two cases will be considered first because they allow analytical solutions of the nonlinear equation system (9.73).

**Equal Initial Values.** For the case $y_{20} = y_{10}$, Eq. (9.74) implies $y_2 / y_1 = y_{20} / y_{10} = 1$, which means $y_2 = y_1$. According to Eq. (9.73a), $y_1$ is then determined by

$$\frac{dy_1}{dt} = y_1\left(1 - (1+d)\, y_1\right). \tag{9.75}$$

This equation is a logistic equation. In particular, this equation corresponds to the logistic equation (7.88) by setting $L = 0$, $\tau = 1$, and $K = 1/(1+d)$. According to the solution (7.101) of the logistic equation, the solution of Eq. (9.75) is given by

$$y_1 = \frac{1/(1+d)}{1 - \left(1 - \dfrac{1/(1+d)}{y_{10}}\right)e^{-t}}. \tag{9.76}$$

The equilibrium solution of Eq. (9.76) is given by $Y_1 = 1/(1+d)$. According to $y_2 = y_1$, the equilibrium point for this case is given by

$$(Y_1, Y_2) = \left( \frac{1}{1+d}, \frac{1}{1+d} \right). \tag{9.77}$$

**Equal Competition.** We also have a proportionality $y_2/y_1 = y_{20}/y_{10}$ for $d = 1$ where we have an equal competition (because the parenthesis terms in Eqs. (9.73) are equal). According to Eq. (9.73a), $y_1$ is then determined by the equation

$$\frac{dy_1}{dt} = y_1 \left( 1 - \left[ 1 + \frac{y_{20}}{y_{10}} \right] y_1 \right). \tag{9.78}$$

The logistic equation (7.88) corresponds to the latter equation if $L = 0$, $\tau = 1$, and $K = 1/(1+y_{20}/y_{10})$. By using the solution (7.101) of the logistic equation, we find the solution of Eq. (9.78) to be given by

$$y_1 = \frac{1/(1+y_{20}/y_{10})}{1 - \left( 1 - \dfrac{1/(1+y_{20}/y_{10})}{y_{10}} \right) e^{-t}}. \tag{9.79}$$

The equilibrium solution that is implied by this expression is $Y_1 = 1/(1+y_{20}/y_{10})$. According to $y_2 = y_1\, y_{20}/y_{10}$, the equilibrium point for this case is

$$(Y_1, Y_2) = \left( \frac{1}{1+y_{20}/y_{10}}, \frac{y_{20}/y_{10}}{1+y_{20}/y_{10}} \right) = \left( \frac{y_{10}}{y_{10}+y_{20}}, \frac{y_{20}}{y_{10}+y_{20}} \right). \tag{9.80}$$

**Linear Stability Analysis.** We have to use linear stability analysis to study the behavior of the nonlinear equation system (9.73) for the cases of unequal initial values and $d \neq 1$. From Eqs. (9.62), the equilibrium values of this model are

$$\begin{aligned} &(Y_1, Y_2) = (0,0), \qquad (Y_1, Y_2) = (0,1), \qquad (Y_1, Y_2) = (1,0), \\ &(Y_1, Y_2) = \left( \frac{1-d}{1-d^2}, \frac{1-d}{1-d^2} \right) = \left( \frac{1}{1+d}, \frac{1}{1+d} \right). \end{aligned} \tag{9.81}$$

Therefore, there are four potential equilibrium states: both species will disappear, only one of the species will survive, or there is a coexistence of both species. For a non-negative $d$ we find that the $Y_1$ and $Y_2$ values are bounded by zero and one, this means $0 \le Y_1 \le 1$ and $0 \le Y_2 \le 1$. In the following, we will analyze the solution behavior in the vicinity of the four equilibrium points (9.81) by making use of the linear stability analysis approach presented in Sect. 9.2.

• $(Y_1, Y_2) = (0, 0)$: the linear equation system (9.69) reads for this case

$$\frac{d}{dt} \begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix}. \tag{9.82}$$

According to Eqs. (9.12), we have for this case the eigenvalues

$$r_1 = 1, \qquad r_2 = 1, \tag{9.83}$$

Thus, the solution is unstable at the equilibrium point $(Y_1, Y_2) = (0, 0)$.

• $(Y_1, Y_2) = (0, 1)$: the linear equation system (9.69) reads for this case

$$\frac{d}{dt}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix} = \begin{pmatrix} 1-d & 0 \\ -d & -1 \end{pmatrix}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix}. \tag{9.84}$$

The eigenvalues are given by

$$r_1 = -\frac{d}{2} + \left|1 - \frac{d}{2}\right|, \qquad r_2 = -\frac{d}{2} - \left|1 - \frac{d}{2}\right|. \tag{9.85}$$

For the case $1 - d/2 > 0$, the eigenvalues are given by $r_1 = 1 - d$ and $r_2 = -1$. For the case $1 - d/2 < 0$, the eigenvalues are given by $r_1 = -1$ and $r_2 = 1 - d$. It is up to us which eigenvalue we called $r_1$ and $r_2$. Thus, we can use

$$r_1 = 1 - d, \qquad r_2 = -1. \tag{9.86}$$

Depending on the value of $d$, the eigenvalue $r_1$ can be negative or positive. Thus, the solution can be asymptotically stable or unstable at this equilibrium point.

• $(Y_1, Y_2) = (1, 0)$: the linear equation system (9.69) reads for this case

$$\frac{d}{dt}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix} = \begin{pmatrix} -1 & -d \\ 0 & 1-d \end{pmatrix}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix}. \tag{9.87}$$

The eigenvalues are given by

$$r_1 = -\frac{d}{2} + \left|1 - \frac{d}{2}\right|, \qquad r_2 = -\frac{d}{2} - \left|1 - \frac{d}{2}\right|, \tag{9.88}$$

That are the same eigenvalues as found for $(Y_1, Y_2) = (0, 1)$. Correspondingly, the eigenvalues are again given by

$$r_1 = 1 - d, \qquad r_2 = -1, \tag{9.89}$$

i.e., the system behavior is the same as in the vicinity of $(Y_1, Y_2) = (0, 1)$.

• $(Y_1, Y_2) = (Y, Y)$, where $Y = 1/(1 + d)$: the equation system (9.69) now reads

$$\frac{d}{dt}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix} = \begin{pmatrix} 1-(2+d)Y & -dY \\ -dY & 1-(2+d)Y \end{pmatrix}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix}. \tag{9.90}$$

The eigenvalues are given by

$$\begin{aligned} r_1 &= 1-(2+d)Y + dY = 1 - 2Y = \frac{1+d-2}{1+d} = \frac{-1+d}{1+d}, \\ r_2 &= 1-(2+d)Y - dY = 1 - 2(1+d)Y = 1 - 2 = -1. \end{aligned} \tag{9.91}$$

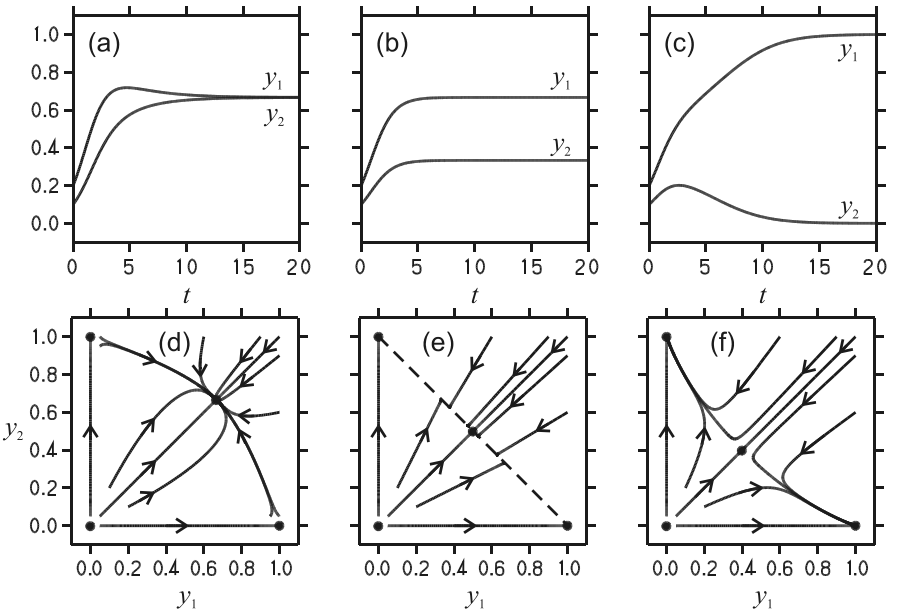



**Fig. 9.3.** The competition of two species $y_1$ and $y_2$ for food. The solutions $y_1$ and $y_2$ of the equation system (9.73) are shown in (a), (b), and (c) as function of time $t$ for $d$ = (0.5, 1, 1.5), respectively. The initial conditions $y_{10}$ = 0.2 and $y_{20}$ = 0.1 are used. The phase plane evolution of $y_1$ and $y_2$ is shown for several initial conditions in (d), (e), and (f) for the cases $d$ = (0.5, 1, 1.5), respectively. The *dashed line* in (e) is not a realizable trajectory: is only gives an orientation regarding the equilibrium points for this case.

As given for the equilibrium point $(Y_1, Y_2) = (0, 1)$, the eigenvalue $r_1$ can be negative or positive depending on the value of $d$. Consequently, the solution can be asymptotically stable or unstable at this equilibrium point.

**Illustration.** We will assume that there are two populations initially such that the initial values $y_{10}$ and $y_{20}$ are nonzero. Then, the equilibrium point (0, 0), which is characterized by two positive eigenvalues, can never be realized. Hence, it is impossible that both populations disappear. For equal initial values $y_{20} = y_{10}$ we find the solution (9.76) for $y_1$ and we have $y_2 = y_1$. The equilibrium solution is $(Y_1, Y_2) = (1 / (1 + d), 1 / (1 + d))$. For unequal initial values we find features that are illustrated in Fig. 9.3 for $d$ = (0.5, 1, 1.5) and several initial conditions. For a relatively weak competition ($d < 1$), we find the development of a coexistence $(Y_1, Y_2) = (1 / (1 + d), 1 / (1 + d))$ between both species. This result agrees with the conclusions of linear stability analysis: the equilibrium points (0, 1) and (1, 0) are

characterized by one positive eigenvalue, whereas the coexistence point ($Y$, $Y$) is characterized by two negative eigenvalues. The equal competition case with $d = 1$ does still allow a coexistence of species, but the initial values matter in this case: the species with the higher initial value will have a higher equilibrium value (see Fig. 9.3b). The solution for $y_1$ is given for this case by the logistic function (9.79), and we have $y_2 = y_1\, y_{20} / y_{10}$. The equilibrium point is given by Eq. (9.80). For a relatively strong competition ($d > 1$) we find that one species disappears whereas the other species achieves a maximum value. In particular, we find that the species with the higher initial value will survive. This observation is also supported by linear stability analysis: the coexistence point ($Y$, $Y$) has one positive eigenvalue, and the equilibrium points (0, 1) and (1, 0) have two negative eigenvalues.

### *9.3.3 Predator-Prey Interaction*

**Model Considered.** The predator-prey equations (9.72) will be also analyzed by considering more specific equations. In correspondence to Eq. (9.73) we apply $a_1 = 1$ and $a_2 = -1$. We assume an equal amount of interaction by setting $c_1 = -4$ and $c_2 = 4$. In addition, we assume that $b_1 = -e$, where $e$ is a non-negative parameter. With these assumptions, the equation system considered is given by

$$\frac{dy_1}{dt} = y_1\left(1 - e\, y_1 - 4 y_2\right), \tag{9.92a}$$

$$\frac{dy_2}{dt} = y_2\left(-1 + 4 y_1\right). \tag{9.92b}$$

The specific relevance of $e$ variations can be seen by considering

$$\frac{dy_2}{dy_1} = \frac{dy_2/dt}{dy_1/dt} = \frac{y_2}{y_1}\frac{-1 + 4 y_1}{1 - e\, y_1 - 4 y_2}. \tag{9.93}$$

This equation is a separable equation for $e = 0$, which means that this equation can be solved.

**Zero Self-Limitation.** First, let us find the analytical solution to the nonlinear equation system (9.92) for the case $e = 0$. Relation (9.93) can be written then

$$\frac{1 - 4 y_2}{y_2} dy_2 - \frac{-1 + 4 y_1}{y_1} dy_1 = 0. \tag{9.94}$$

The integration of both sides provides

$$\ln |\, y_2 \,| - 4 y_2 + \ln |\, y_1 \,| - 4 y_1 = C. \tag{9.95}$$

Here, $C$ is a constant that is determined through the initial conditions (by setting $t = 0$ on the left-hand side). Unfortunately, it is impossible to use this relation for the calculation of $y_2$ as an explicit function of $y_1$. The relevance of Eq. (9.95) is that this relation describes a closed curve (see, for example, the illustration of this case in Fig. 9.4d). The existence of a closed curve means that we have a stable solution for $e = 0$. Hence, nonzero $e$ values describe deviations from a stable state. Interestingly, a closed curve is also found for any other parameter values than those used in Eq. (9.92), provided these parameters have the same signs and $e = 0$.

**Linear Stability Analysis.** We have to use again linear stability analysis to understand the behavior of the nonlinear equation system (9.92) for the case $e \neq 0$. According to Eq. (9.62), there are three potential equilibrium points for this system (the equilibrium point $(0, -a_2/b_2)$ in Eq. (9.62) cannot be realized),

$$(Y_1, Y_2) = (0,0), \quad (Y_1, Y_2) = \left(\frac{1}{e}, 0\right), \quad (Y_1, Y_2) = \left(\frac{4}{16}, \frac{4-e}{16}\right) = \left(\frac{1}{4}, \frac{1}{4}\left(1 - \frac{e}{4}\right)\right). \tag{9.96}$$

Correspondingly, it is possible that both species will be extinct, or only the prey survives, or there is a coexistence of both species. The equilibrium values $Y_1$ and $Y_2$ are positive if $e \leq 4$. The setting $e = 4$ does recover the second equilibrium point (1 / 4, 0). For simplicity, we do not consider this case $e = 4$, this means we consider variations $0 \leq e < 4$. The solution behavior of Eqs. (9.92) in the vicinity of the three equilibrium points $(Y_1, Y_2)$ reveals the following features:

- $(Y_1, Y_2) = (0, 0)$: the linear equation system (9.69) reads for this case

$$\frac{d}{dt}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix}. \tag{9.97}$$

According to Eqs. (9.12), the eigenvalues are given by

$$r_1 = 1, \qquad r_2 = -1, \tag{9.98}$$

Correspondingly, the solution is unstable at the equilibrium point $(Y_1, Y_2) = (0, 0)$.

- $(Y_1, Y_2) = (1/e, 0)$: the linear equation system (9.69) reads for this case

$$\frac{d}{dt}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix} = \begin{pmatrix} -1 & -\frac{4}{e} \\ 0 & -1 + \frac{4}{e} \end{pmatrix}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix}. \tag{9.99}$$

The eigenvalues are provided by

$$r_1 = -1 + \frac{2}{e} + \frac{2}{e} = -1 + \frac{4}{e}, \qquad r_2 = -1 + \frac{2}{e} - \frac{2}{e} = -1. \tag{9.100}$$

The eigenvalue $r_1$ is positive due to the condition $0 \leq e < 4$. Thus, the solution is unstable at the equilibrium point $(Y_1, Y_2) = (1/e, 0)$.

• $(Y_1, Y_2) = (1/4, [1 - e/4]/4)$: the equation system (9.69) now reads

$$\frac{d}{dt}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix} = \begin{pmatrix} 1 - 2eY_1 - 4Y_2 & -4Y_1 \\ 4Y_2 & -1 + 4Y_1 \end{pmatrix}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix}. \tag{9.101}$$

By adopting the definitions of $Y_1$ and $Y_2$ the latter equation system can be written in the following way,

$$\frac{d}{dt}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix} = \begin{pmatrix} 1 - \frac{e}{2} - \left(1 - \frac{e}{4}\right) & -1 \\ 1 - \frac{e}{4} & 0 \end{pmatrix}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix} = \begin{pmatrix} -\frac{e}{4} & -1 \\ 1 - \frac{e}{4} & 0 \end{pmatrix}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \end{pmatrix}. \tag{9.102}$$

Therefore, the eigenvalues are given by

$$\begin{aligned} r_1 &= -\frac{e}{8} + \frac{1}{2}\sqrt{\frac{e^2}{16} - 4\left(1 - \frac{e}{4}\right)} = -\frac{1}{2}\left(\frac{e}{4} - \sqrt{\left(\frac{e}{4} + 2\right)^2 - 8}\right), \\ r_2 &= -\frac{e}{8} - \frac{1}{2}\sqrt{\frac{e^2}{16} - 4\left(1 - \frac{e}{4}\right)} = -\frac{1}{2}\left(\frac{e}{4} + \sqrt{\left(\frac{e}{4} + 2\right)^2 - 8}\right). \end{aligned} \tag{9.103}$$

Depending on the $e$ variation, $0 \le e/4 < 1$, there are three cases of eigenvalues. A first case is given for $e = 0$, which means that we have two complex eigenvalues with zero real parts. As discussed above, this case corresponds to a stable solution. For a nonzero $e$, we have to distinguish cases for which the square root is real and imaginary. The square root becomes zero for $e/4 = \pm 8^{1/2} - 2$. Due to the variation $0 \le e/4 < 1$ considered, only the value $e/4 = 8^{1/2} - 2 = 0.8284$ can be realized. Correspondingly, we may have two cases in addition to the case $e = 0$. For the case $0 < e/4 < 0.8284$, we have two complex eigenvalues with negative real part. Thus, the solution is asymptotically stable. For the case $0.8284 < e/4 < 1$ we have a real square root. The eigenvalues $r_1$ and $r_2$ are always negative for this case. Hence, the solution is again asymptotically stable.

**Illustration.** An illustration of solutions of the equation system (9.92) is given in Fig. 9.4 for the cases $e/4 = (0, 0.6, 0.9)$. We consider nonzero initial populations densities $y_{10}$ and $y_{20}$. This assumption implies that the equilibrium solutions (0, 0) and $(1/e, 0)$, which are both characterized by one positive eigenvalue, can never be realized. Consequently, no population will disappear, which means that there will be a coexistence between both populations. For a zero self-limitation (Fig. 9.4a, d), we observe cyclic variations of the predator and prey populations: a decrease (increase) of prey leads after a delay time to an increase (decrease) of predators. An equilibrium state cannot be established in this way. In the phase plane, the latter behavior corresponds to a closed curve that surrounds the center.

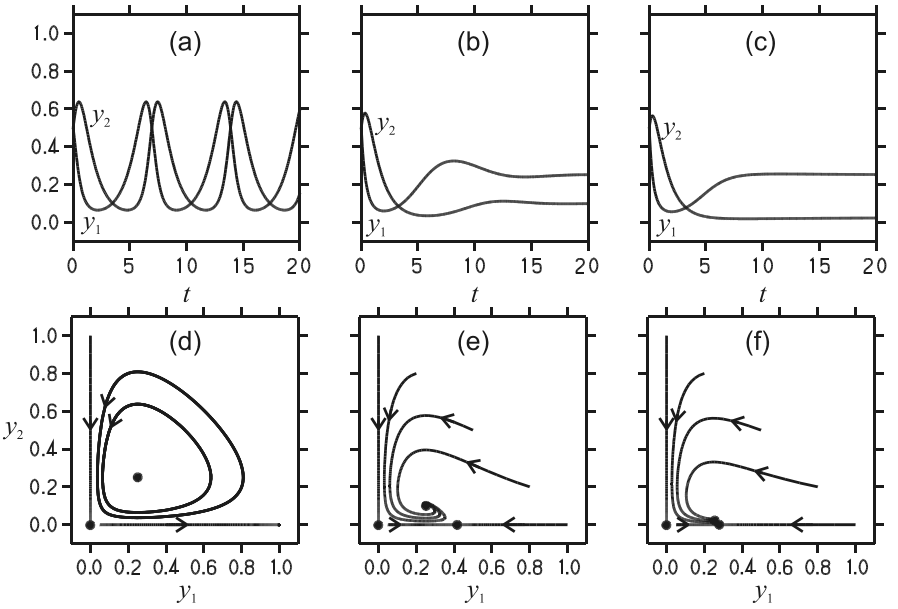



**Fig. 9.4.** Interactions between a predator $y_2$ and prey $y_1$. The solutions $y_1$ and $y_2$ of the equations (9.92) are shown in (a), (b), and (c) as function of time $t$ for the cases $e / 4 = (0, 0.6, 0.9)$, respectively, where $y_{10} = y_{20} = 0.5$. The corresponding phase plane evolution of $y_1$ and $y_2$ is shown for several initial conditions in (d), (e), and (f), where $e / 4 = (0, 0.6, 0.9)$, respectively. The *dots* show the equilibrium points.

The $y_1$-$y_2$ curves follow Eq. (9.95), and the center is located at (1/4, 1/4), see Eq. (9.96). For a nonzero but relatively weak self-limitation $0 < e / 4 < 0.8284$ (see Fig. 9.4b, e), the prey curve shows oscillations that are damped out due to the self-limitation (which appears as a sink term in Eq. (9.92a)). Due to the coupling with $y_1$, the predator curve also shows damped oscillations. The damping implies in the phase plane an asymptotically stable solution. For the case of a relatively strong self-limitation $0.8284 < e/4 < 1$ (see Fig. 9.4c, f), the damping does not allow oscillations anymore. After the first minimum (maximum), $y_1$ ($y_2$) realizes the equilibrium value. It is interesting to see that the increasing damping reduces the $Y_2$ coordinate of the coexistence point $(Y_1, Y_2) = (1/4, [1 - e / 4] / 4)$, whereas the $Y_1$ coordinate is unaffected. The latter fact is a consequence of Eq. (9.92b), which fixes the stationary value $Y_1 = 1/4$. A modification of this equation (e.g., by the addition of a positive term proportional to $y_2$ in the parenthesis term) would lead to different features.

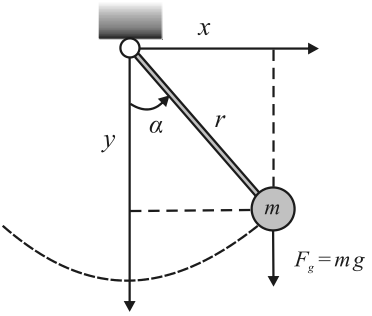



**Fig. 9.5.** An illustration of a pendulum.

## 9.4 Mechanical Motions: The Pendulum

Let us consider next evolution principles for vector processes in mechanics. In continuation of the explanation of laws for mechanical processes in Chap. 7 we will focus the discussion in this chapter on the application of Newton's Laws of Motion. In particular, we will extend the discussion of one-dimensional harmonic oscillator motions in Chap. 7 by considering now the motions of a pendulum. The results obtained in this way were used in Chap. 3 regarding the discussion of the measurement of time.

### *9.4.1 Pendulum Equations*

**Newton's Laws of Motion.** Contrary to population ecology we have a sound mathematical basis for the modeling of mechanical processes given by Newton's Laws of Motion, which were discussed in Chap. 7. Mechanical motions of macroscopic bodies that move with velocities much smaller than the speed of light can be described by Newton's Second Law given by Eq. (7.33),

$$\frac{d^2\boldsymbol{x}}{dt^2} = \frac{\boldsymbol{F}}{m}. \tag{9.104}$$

Here, $\boldsymbol{x} = (x_1, x_2, x_3)$ is the position vector of any body, $\boldsymbol{F} = (F_1, F_2, F_3)$ is the force acting on the body, and $m$ is the mass of the body. The use of this equation for the calculation of pendulum motions will be demonstrated in the following.

**Undamped Pendulum Equation.** An illustration of the pendulum considered is given in Fig. 9.5. A mass $m$ is attached to one end of a rigid, but weightless, supported rod of length $r$. The rod is free to rotate in one plane. The angle $\alpha(t)$ is the angle of displacement from the vertical. The force that drives the pendulum is given by the gravity force $F_g = m\,g$. Here, $g$ denotes the gravity acceleration. Other

forces are not involved regarding the undamped pendulum motion. The $x$-$y$ coordinate system applied is shown in Fig. 9.5. In correspondence to the analysis of the spring-mass system we assume that the downward direction $y$ is the positive direction. The equations that govern the motion of the pendulum (the changes of the $x(t)$ and $y(t)$ coordinates of the pendulum) are given by Newton's Second Law (9.104).

$$\frac{d^2x}{dt^2} = 0, \tag{9.105a}$$

$$\frac{d^2y}{dt^2} = \frac{F_g}{m} = g. \tag{9.105b}$$

To take advantage of the fact that the pendulum moves along a circle with constant radius $r$, it is helpful to switch to polar coordinates given by the radius $r$ and the angle of displacement $\alpha$. According to the illustration in Fig. 9.5, the relations that relate $(x, y)$ and $(r, \alpha)$ are given by

$$\sin\alpha = \frac{x}{r}, \qquad \cos\alpha = \frac{y}{r}. \tag{9.106}$$

The first-order and second-order derivatives of $x(t)$ and $y(t)$ that are implied by these relations are given by ($r$ is constant)

$$\frac{dx}{dt} = r\cos\alpha\frac{d\alpha}{dt}, \qquad \frac{d^2x}{dt^2} = -r\sin\alpha\left(\frac{d\alpha}{dt}\right)^2 + r\cos\alpha\frac{d^2\alpha}{dt^2}, \tag{9.107a}$$

$$\frac{dy}{dt} = -r\sin\alpha\frac{d\alpha}{dt}, \qquad \frac{d^2y}{dt^2} = -r\cos\alpha\left(\frac{d\alpha}{dt}\right)^2 - r\sin\alpha\frac{d^2\alpha}{dt^2}. \tag{9.107b}$$

The use of these relations in Newton's Second Law equations (9.105) then implies

$$-r\sin\alpha\left(\frac{d\alpha}{dt}\right)^2 + r\cos\alpha\frac{d^2\alpha}{dt^2} = 0, \tag{9.108a}$$

$$-r\cos\alpha\left(\frac{d\alpha}{dt}\right)^2 - r\sin\alpha\frac{d^2\alpha}{dt^2} = g. \tag{9.108b}$$

Equation (9.108a) can be used to replace the quadratic first-order derivative by the second-order derivative of $\alpha$,

$$\left(\frac{d\alpha}{dt}\right)^2 = \frac{\cos\alpha}{\sin\alpha}\frac{d^2\alpha}{dt^2}. \tag{9.109}$$

The use of this relation in Eq. (9.108b) leads then to

$$\left[\cos\alpha\frac{\cos\alpha}{\sin\alpha} + \sin\alpha\right]\frac{d^2\alpha}{dt^2} = \frac{\cos^2\alpha + \sin^2\alpha}{\sin\alpha}\frac{d^2\alpha}{dt^2} = \frac{1}{\sin\alpha}\frac{d^2\alpha}{dt^2} = -\frac{g}{r}, \tag{9.110}$$

where the Pythagorean identity was applied. Hence, the equation of motion for the undamped pendulum reads

$$\frac{d^2\alpha}{dt^2} = -\frac{g}{r}\sin\alpha. \tag{9.111}$$

**Damped Pendulum Equation.** In general, the pendulum will be also affected by a damping force, which reduces the pendulum velocity due to the air resistance. In correspondence to the analysis of the spring-mass system we assume that this damping force is proportional to the pendulum velocity $d\alpha/dt$. Hence, we extend Eq. (9.111) in the following way,

$$\frac{d^2\alpha}{dt^2} = -\frac{1}{\tau}\frac{d\alpha}{dt} - \frac{g}{r}\sin\alpha. \tag{9.112}$$

For the characteristic damping time scale we use Stokes' Law (see Sect. 3.3.3),

$$\tau = \frac{m}{6\pi\mu r_P}. \tag{9.113}$$

Here, $\mu$ refers to the dynamic viscosity, and $r_P$ is the radius of the spherical mass. The damping contribution appears in Eq. (9.112) with a negative sign because this term reduces the pendulum velocity $d\alpha/dt$ (the damping term implies that $d\alpha/dt$ becomes smaller for a positive $d\alpha/dt$). Similar as in the discussion of damping in the spring-mass system, the structure of the damping term applied here does only represent one reasonable assumption among several possible choices. The pendulum equation that results from the use of Eq. (9.113) then reads

$$\frac{d^2\alpha}{dt^2} + \frac{6\pi\mu r_P}{m}\frac{d\alpha}{dt} + \frac{g}{r}\sin\alpha = 0. \tag{9.114}$$

**Normalized Damped Pendulum Equation.** The use of nondimensional variables is helpful because the number of model parameters involved in the equation can be reduced. We introduce the nondimensional time $t_* = t/(r/g)^{1/2}$, such that Eq. (9.114) reads

$$\frac{d^2\alpha}{dt_*^2} + \frac{6\pi\mu r_P}{m}\sqrt{\frac{r}{g}}\frac{d\alpha}{dt_*} + \sin\alpha = 0. \tag{9.115}$$

By introducing the nondimensional dynamic viscosity $\mu_* = \mu\,(r^3/g)^{1/2}/m$, which can be seen as an inverse Reynolds number, this equation can be written

$$\frac{d^2\alpha}{dt_*^2} + 6\pi\mu_*\frac{r_P}{r}\frac{d\alpha}{dt_*} + \sin\alpha = 0. \tag{9.116}$$

We introduce the nondimensional variable $d = 6\pi r_P/r$ to simplify the writing of this equation. Then, the damped pendulum equation, which does now only depend

on one parameter (the product $d\,\mu_*$), is given by

$$\frac{d^2\alpha}{dt_*^2} + d\,\mu_* \frac{d\alpha}{dt_*} + \sin\alpha = 0. \tag{9.117}$$

**Nonlinear Equation System.** The second-order differential equation (9.117) can also be represented as an equation system. We set $y_1 = \alpha$ and $y_2 = d\alpha/dt_*$. The differentiation of $y_1$ and $y_2$ provides then

$$\frac{dy_1}{dt_*} = \frac{d\alpha}{dt_*} = y_2, \tag{9.118a}$$

$$\frac{dy_2}{dt_*} = \frac{d^2\alpha}{dt_*^2} = -d\,\mu_* \frac{d\alpha}{dt_*} - \sin\alpha = -d\,\mu_* y_2 - \sin y_1. \tag{9.118b}$$

The initial values for $y_1$ and $y_2$ are given by $y_{10} = \alpha(0)$ and $y_{20} = d\alpha/dt_*(0)$.

## *9.4.2 Linear Stability Analysis*

Let us analyze the nonlinear equation system (9.118) by adopting the linear stability analysis approach described in Sect. 9.2.3.

**Equilibrium Points.** First, we have to determine the equilibrium solutions of Eq. (9.118). Such equilibrium solutions have to satisfy the equations

$$0 = y_2, \tag{9.119a}$$

$$0 = -d\,\mu_* y_2 - \sin y_1. \tag{9.119b}$$

These two equations are solved by $y_1 = \pm\, n\,\pi$ and $y_2 = 0$, where $n = 0, 1, 2, \ldots$. However, there is no need to consider all these equilibrium points. We do only have to consider the two physical equilibrium solutions (0, 0) and ($\pi$, 0). Due to physical reasons we expect that the first equilibrium point (0, 0) is asymptotically stable and the second equilibrium point ($\pi$, 0) is unstable.

**First Equilibrium Point: Linear Stability Analysis.** According to Eq. (9.69), the linear equation system that describes the pendulum motion close to the first equilibrium point (0, 0) reads

$$\frac{d}{dt_*}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & -d\,\mu_* \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}. \tag{9.120}$$

By differentiating $dy_1/dt_* = y_2$ and using $dy_2/dt_* = -y_1 - d\,\mu_* y_2$, we find in terms of the original variables $y_1 = \alpha$ and $y_2 = d\alpha/dt_*$ the equation

$$\frac{d^2\alpha}{dt_*^2} = -\alpha - d\,\mu_* \frac{d\alpha}{dt_*}. \tag{9.121}$$

Hence, the analysis of this case corresponds to the consideration of the linearized pendulum equation (9.117) where sin $\alpha$ is approximated by $\alpha$, which is justified for sufficiently small initial angles of displacement. The equation system (9.120) represents a specific case of the linear equation system (9.2). The eigenvalues of Eq. (9.120) are, therefore, given by Eq. (9.12),

$$r_1 = r_S + r_D, \qquad r_2 = r_S - r_D, \tag{9.122}$$

where $r_S$ and $r_D$ are given by

$$r_S = -d\,\mu_*/2, \qquad r_D = \sqrt{(d\,\mu_*/2)^2 - 1}. \tag{9.123}$$

The effect of damping is relatively small in general. Therefore, we may assume that $d\,\mu_* < 2$. For this case we find $r_D = i\,r_{D^*}$, where the real number $r_{D^*}$ is

$$r_{D^*} = \sqrt{1 - (d\,\mu_*/2)^2}. \tag{9.124}$$

Both eigenvalues have a negative real part $r_S$ if there is a nonzero damping, i.e., $d\,\mu_* \neq 0$. The discussion at the end of Sect. 9.2.2 showed that such a system is asymptotically stable. The solution to the linear equation system (9.120) can be found by making use of the fact that Eq. (9.121) represents a specific case of the homogeneous linear second-order differential equation (7.45). According to Eq. (7.69), the solution of Eq. (9.121) is then given by

$$\begin{aligned}\alpha(t_*) &= e^{r_S t_*}\left\{\frac{\sin(r_{D^*} t_*)}{r_{D^*}}\alpha'_0 + \left[\cos(r_{D^*} t_*) - \frac{\sin(r_{D^*} t_*)}{r_{D^*}} r_S\right]\alpha_0\right\} \\ &= \alpha_0\, e^{r_S t_*}\left\{\cos(r_{D^*} t_*) + \frac{\alpha'_0 - r_S\alpha_0}{r_{D^*}\alpha_0}\sin(r_{D^*} t_*)\right\}.\end{aligned} \tag{9.125}$$

Here, $\alpha_0$ refers to the initial angle of displacement $\alpha(0)$, and $\alpha'_0$ is the initial value of $d\alpha/dt_*$. The latter relation can be rewritten by defining an angle $\delta$ by

$$\tan\delta = \frac{\alpha'_0 - r_S\alpha_0}{r_{D^*}\alpha_0}. \tag{9.126}$$

The use of this relation then enables the following rewriting of Eq. (9.125),

$$\begin{aligned}\alpha(t_*) &= \left[\cos(r_{D^*} t_*) + \sin(r_{D^*} t_*)\tan\delta\right] e^{r_S t_*}\,\alpha_0 \\ &= \frac{\cos(r_{D^*} t_*)\cos\delta + \sin(r_{D^*} t_*)\sin\delta}{\cos\delta} e^{r_S t_*}\,\alpha_0 = \frac{\cos(r_{D^*} t_* - \delta)}{\cos\delta} e^{r_S t_*}\,\alpha_0.\end{aligned} \tag{9.127}$$

By replacing $r_S$ and $r_{D^*}$ according to their definitions (9.123) and (9.124) we find

$$\alpha(t_*) = \frac{\cos\left(\sqrt{1 - (d\,\mu_*/2)^2}\; t_* - \delta\right)}{\cos\delta} e^{-d\,\mu_* t_*/2}\,\alpha_0, \tag{9.128}$$

where the angle $\delta$ is given by

$$\delta = \arctan\left( \frac{\alpha'_0 / \alpha_0 + d\,\mu_* / 2}{\sqrt{1 - (d\,\mu_* / 2)^2}} \right). \tag{9.129}$$

The solution of the equations (9.120) is then given by $y_1 = \alpha$ and $y_2 = d\alpha / dt_*$.

**Second Equilibrium Point: Linear Stability Analysis.** At the second equilibrium point ($\pi$, 0) Eqs. (9.69) imply the linear equation system

$$\frac{d}{dt_*}\begin{pmatrix} y_1 - \pi \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -d\,\mu_* \end{pmatrix}\begin{pmatrix} y_1 - \pi \\ y_2 \end{pmatrix}. \tag{9.130}$$

Here, $dy_1 / dt_*$ was replaced by $d(y_1 - \pi) / dt_*$. To solve this equation system we use the equivalent second-order equation

$$\frac{d^2(\alpha - \pi)}{dt_*^{\,2}} = \alpha - \pi - d\,\mu_* \frac{d(\alpha - \pi)}{dt_*}, \tag{9.131}$$

where $y_1 = \alpha$ and $y_2 = d\alpha / dt_*$ are used. The latter equation can be obtained in the same way as Eq. (9.121). The eigenvalues can be written

$$r_1 = -d\,\mu_* / 2 + \sqrt{1 + (d\,\mu_* / 2)^2}, \qquad r_2 = -d\,\mu_* / 2 - \sqrt{1 + (d\,\mu_* / 2)^2}. \tag{9.132}$$

The eigenvalue $r_1$ is always positive (also for zero damping), i.e., the solution near the second equilibrium point is unstable. According to Eq. (7.57), the solution of Eq. (9.131) is given for this case of two unequal real eigenvalues by

$$\alpha - \pi = \frac{\alpha'_0 - r_2(\alpha_0 - \pi)}{r_1 - r_2} e^{r_1 t_*} - \frac{\alpha'_0 - r_1(\alpha_0 - \pi)}{r_1 - r_2} e^{r_2 t_*}. \tag{9.133}$$

The solutions of Eqs. (9.130) follow then from $y_1 = \alpha$ and $y_2 = d\alpha / dt_*$. Regarding the discussion of the evolution in the $y_1$-$y_2$ phase plane below it is interesting to consider the consequences of setting $\alpha'_0 = r_2\,(\alpha_0 - \pi)$, such that the first term in Eq. (9.133) disappears. By differentiating the resulting expression we find

$$\frac{d\alpha}{dt_*} = -r_2 \frac{\alpha'_0 - r_1(\alpha_0 - \pi)}{r_1 - r_2} e^{r_2 t_*} = r_2(\alpha - \pi), \tag{9.134}$$

which means that $y_2 = r_2\,(y_1 - \pi)$ for all $t$. Similarly, we find for $\alpha'_0 = r_1\,(\alpha_0 - \pi)$

$$\frac{d\alpha}{dt_*} = r_1 \frac{\alpha'_0 - r_2(\alpha_0 - \pi)}{r_1 - r_2} e^{r_1 t_*} = r_1(\alpha - \pi), \tag{9.135}$$

which means that $y_2 = r_1\,(y_1 - \pi)$ for all $t$. Hence, the $y_1$-$y_2$ phase plane figure will involve the two linear functions $y_2 = r_2\,(y_1 - \pi)$ and $y_2 = r_1\,(y_1 - \pi)$ in the vicinity of the point ($\pi$, 0) for both the damped and undamped pendulum.

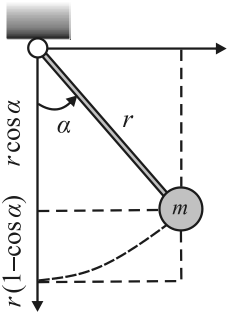



**Fig. 9.6.** An illustration regarding the calculation of the potential energy of a pendulum. The pendulum is lifted above its minimal position by the distance $r(1 - \cos\alpha)$.

### *9.4.3 Nonlinear Stability Analysis*

**Lyapunov's Second Method.** The use of linear stability analysis leads to very helpful conclusions, but the suitability of the assumption of linear processes in the vicinity of equilibrium points is often not very clear. A nice way to overcome this problem by the analysis of the nonlinear equation system (9.118) was developed by Lyapunov. This approach, which will be presented in the following, is known as Lyapunov's second method (Lyapunov's first method refers to the method of linearization of a nonlinear equation along an orbit). The basic idea of Lyapunov's approach is the consideration of orbits (i.e., trajectories in the phase plane) that are characterized by decreasing values of a non-negative function (which is called the Lyapunov function). The trajectory and its Lyapunov function will change until the Lyapunov function reaches the value zero. The position of the trajectory in the phase plane at which the Lyapunov function is equal to zero characterizes an equilibrium point. Hence, the asymptotical stability of nonlinear equation systems can be shown by proving the existence of a Lyapunov function that decreases to zero.

**Pendulum Lyapunov Function.** The most natural choice for the pendulum Lyapunov function is the total energy $E$ defined by

$$E = m\,g\,r(1-\cos\alpha) + \frac{1}{2}mr^2\left(\frac{d\alpha}{dt}\right)^2. \tag{9.136}$$

The first contribution represents the potential energy (the work done in lifting the pendulum above its minimal position: see the illustration in Fig. 9.6). The second contribution is the kinetic energy of the pendulum. To simplify the analysis below we use the nondimensional time $t_* = t/(r/g)^{1/2}$, and we introduce the nondimensional energy $E_* = E/(m\,g\,r)$,

$$E_* = 1-\cos\alpha + \frac{1}{2}\left(\frac{d\alpha}{dt_*}\right)^2 = 1-\cos y_1 + \frac{1}{2}{y_2}^2, \tag{9.137}$$

where the definitions $y_1 = \alpha$ and $y_2 = d\alpha / dt_*$ are applied. The total energy $E_*$ has two relevant properties. The first property is that $E$ is non-negative,

$$E_* \geq 0. \tag{9.138}$$

The case $E_* = 0$ can only appear if $y_2 = 0$ and $y_1 = 2\,n\,\pi$, where $n = 0, \pm1, \pm2, \ldots$ (such that $\cos y_1 = 1$). Hence, $E_* = 0$ for all the asymptotically stable equilibrium positions. The second property of $E_*$ is the inequality

$$\frac{dE_*}{dt_*} = \sin y_1 \frac{dy_1}{dt_*} + y_2 \frac{dy_2}{dt_*} = y_2 \sin y_1 - y_2\left(d\,\mu_* y_2 + \sin y_1\right) = -d\,\mu_* {y_2}^2 \leq 0. \tag{9.139}$$

The derivatives of $y_1$ and $y_2$ are replaced here according to Eqs. (9.118).

**Nonlinear Stability Analysis.** By excluding equilibrium points as initial points we find, therefore, the following results of this discussion.

- The damped pendulum motion is characterized by a decreasing energy $E_*$ except for the case that $y_2 = 0$. There are two possibilities to find this case. First, we have $y_2 = 0$ at the equilibrium points $y_1 = \pm\, n\, \pi$ with $n = 0, \pm1, \pm2, \ldots$. However, the equilibrium points can only be realized asymptotically. Second, $y_2 = 0$ at the points on the left and right side at which the pendulum reverses the direction. However, $E_*$ continues to decrease after passing these points. Thus, the damped pendulum motion is characterized by trajectories with $dE_*/dt < 0$ that approach the asymptotically stable equilibrium points with $y_2 = 0$ and $y_1 = 2\,n\,\pi$, where $n = 0, \pm1, \pm2, \ldots$ (because $E_* = 0$ at these points: see Eq. (9.137)). Hence, the other equilibrium points (as $(-\pi, 0)$, $(\pi, 0)$, $(3\pi, 0)$, …) are asymptotically unstable.
- The undamped pendulum motion ($d\,\mu_* = 0$) is characterized by a constant value $E_* > 0$ (asymptotically stable equilibrium positions are not considered as initial points). The curves satisfy the equation

$$1 - \cos y_1 + \frac{1}{2} {y_2}^2 = E_*. \tag{9.140}$$

The type of curve depends on the value of $E_*$. For relatively small $E_*$ values we have relatively small $y_1$ and $y_2$. By approximating $\cos y_1$ by its Taylor series at (0, 0) in the first order of approximation, $\cos y_1 = 1 - {y_1}^2/2$, Eq. (9.140) becomes

$${y_1}^2 + {y_2}^2 = 2E_*. \tag{9.141}$$

This equation describes a circle centered at (0, 0) with radius $(2E_*)^{1/2}$. For larger $E_*$ values we have to consider the curve formula

$$y_2 = \pm\sqrt{2(E_* + \cos y_1 - 1)}, \tag{9.142}$$

which is implied by Eq. (9.140). Closed curves correspond to stable cyclic motions about the equilibrium point. Closed curves must include the possibility

that $y_2 = d\alpha / dt_* = 0$ such that the pendulum can reverse the direction. The function (9.142) shows that the value $y_2 = 0$ can be realized as long as $E_* \leq 2$ (otherwise $E_*$ is always larger than $\cos y_1 - 1$). Therefore, open curves (corresponding to values $y_1 = \alpha$ that do always increase: see the illustration in the next subsection) are found if $E_* > 2$. The case $E_* = 2$ is a specific case that separates closed and open curves. For this case we have

$$y_2 = \pm\sqrt{2(2 + \cos y_1 - 1)}. \tag{9.143}$$

This curve is called the separatrix for the undamped pendulum motion.

### *9.4.4 Pendulum Motions*

Nonlinear pendulum motions are illustrated in Fig. 9.7 as functions of time and in the $y_1$-$y_2$ phase plane. Typical features of these motions will be discussed next and compared to the conclusions of stability theory.

**Undamped Pendulum.** The undamped pendulum is characterized by two sorts of areas: areas that are bounded from below and above by separatrices (indicated by the differently shaded areas in Fig. 9.7c), and the remaining areas. The shaded areas are characterized by closed curves in the $y_1$-$y_2$ phase plane corresponding to cyclic pendulum motions: see the $y_1(t)$ curve for $y_{20} = \pi/3$. The system behavior is very different outside the shaded areas: we have here open curves in the $y_1$-$y_2$ phase plane that are related to a steady increase of $y_1 = \alpha$: see the $y_1$ curves related to $y_{20} = 2\pi/3$ and $y_{20} = 3\pi/2$. The separatrices separate these two behaviors. The separatrices are closed, but the corresponding $y_1(t)$ curve does not show cyclic variations anymore: see the curve $y_1(t)$ that results from $y_{20} = 2^{1/2}$.

**Comparison with Stability Theory.** These observations agree with the consequences of linear and nonlinear stability theory. The local phase plane features are explained by the linear stability theory. The two linear functions $y_2 = r_2 (y_1 - \pi)$ and $y_2 = r_1 (y_1 - \pi)$ are found in the vicinity of the point $(\pi, 0)$: $y_2 = r_1 (y_1 - \pi)$ is the increasing function, and $y_2 = r_2 (y_1 - \pi)$ is the decreasing function. The global phase plane features are explained by the nonlinear stability theory. The curve shapes correspond to the conclusions reported in Sect. 9.4.3 as a consequence of analyzing the Lyapunov function $E_*$. For relatively small $E_*$ values we find circles, the separatrix obtained for $E_* = 2$ is described by Eq. (9.143), and for large values of $E_*$ we find open curves.

**Damped Pendulum.** The phase plane for the damped pendulum is differently organized. All the space is divided into areas that are enclosed by separatrices: the differently shaded areas in Fig. 9.7d are surrounded by other areas that are also enclosed by separatrices. The calculation of separatrices is not as simple as for the

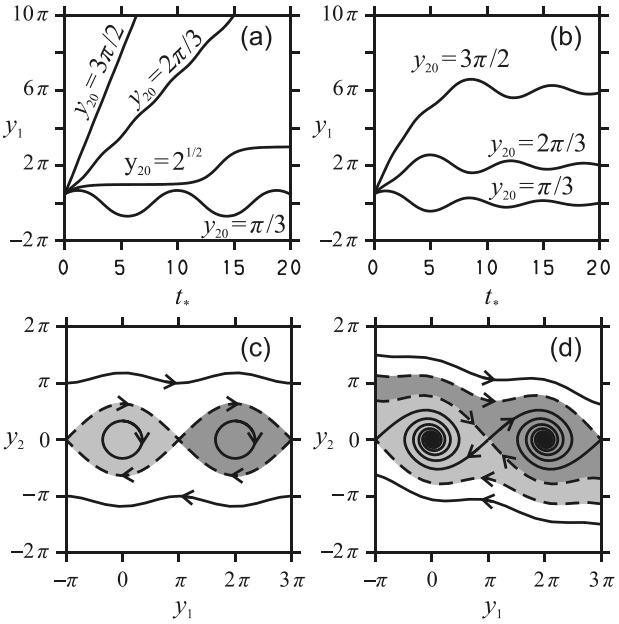



**Fig. 9.7.** Pendulum motions. Solutions $y_1 = \alpha$ of the nonlinear pendulum equation system (9.118) are shown as function of time $t_*$ in (a) and (b) for the undamped ($d\,\mu_* = 0$) and damped pendulum ($d\,\mu_* = 0.2$), respectively. The curves start at $y_{10} = \pi/2$. The initial values $y_{20}$ have the values given in the figures; (c) and (d) illustrate undamped and damped pendulum motion, respectively, for a variety of initial conditions in the $y_1$-$y_2$ phase plane. Separatrices are indicated by *dashed lines*. The differently *shaded* areas indicate areas enclosed by separatrices.

undamped pendulum because the choice of initial values for the trajectories is not obvious. In particular, the lowest separatrix has to be calculated such that it ends in (0, 0). This curve is determined by the initial values ($3\,\pi$, −3.574). The middle separatrix between $\pi$ and $3\,\pi$ can be calculated by the initial vales ($5\,\pi$, −3.574). The other separatrices follow from symmetry conditions: The middle separatrix between $-\pi$ and $\pi$ follows from the initial value ($-3\,\pi$, 3.574), and the highest separatrix follows from the initial value ($3\,\pi$, 3.574). Curves that begin inside the lower shaded area in Fig. 9.7d are attracted by the equilibrium point (0, 0), whereas curves that begin inside the upper shaded area are attracted by the equilibrium point ($2\,\pi$, 0). Correspondingly, curves that begin in other areas are attracted by different asymptotically stable equilibrium points. For example, the lowest solid curve in Fig. 9.7d is attracted by ($-2\,\pi$, 0), and the highest solid line is attracted by ($4\,\pi$, 0).

**Comparison with Stability Theory.** These observations do also agree with the conclusions of stability theory. According to linear stability theory, the linear functions $y_2 = r_2\,(y_1 - \pi)$ and $y_2 = r_1\,(y_1 - \pi)$ are found at $(\pi, 0)$: $y_2 = r_1\,(y_1 - \pi)$ is the increasing function, and $y_2 = r_2\,(y_1 - \pi)$ is the decreasing function. According to nonlinear stability analysis, all trajectories approach the asymptotically stable equilibrium points with $y_2 = 0$ and $y_1 = 2\,n\,\pi$, where $n = 0, \pm 1, \pm 2, \ldots$. Thus, the other equilibrium points $(-\pi, 0)$, $(\pi, 0)$, $(3\,\pi, 0)$, … are asymptotically unstable.

## 9.5 Fluid Dynamics: Lorenz's Weather

Next, let us analyze Lorenz's weather. The latter term refers to a simple model for the explanation of convection, which is relevant to atmospheric motions (the motion of the atmosphere is forced by the latitudinal imbalance of solar heating) and many technical applications (e.g., the design of heat exchangers). The consideration of convection here continues the discussion of simple convection models in Sect. 7.2. The model equations considered represent a simplification of the complicated partial differential equations of fluid dynamics derived in Chap. 10, which are implied by the Newtonian mechanics. A main feature of fluid dynamics equations is that these nonlinear equations generate deterministic chaos (chaotic solutions). It will be shown below that Lorenz's weather model is characterized by the same feature. From a methodological point of view, the equations considered extend the previous analyses by the consideration of three coupled equations.

### *9.5.1 The Lorenz Equations*

**Lorenz's Equations.** Following the studies of Saltzman (1962), Lorenz (1963) suggested the following equations to investigate basic features of convection,

$$\frac{dy_1}{dt} = Pr\,(y_2 - y_1), \tag{9.144a}$$

$$\frac{dy_2}{dt} = y_1(R - y_3) - y_2, \tag{9.144b}$$

$$\frac{dy_3}{dt} = y_1 y_2 - b\,y_3. \tag{9.144c}$$

These equations represent a highly simplified model for (Rayleigh–Bénard) convection. Lorenz's equations can be also seen as a toy model for weather, which can be used for explaining the limitations of long-range weather forecasting (see e.g., Gleick 1987, Lorenz 2006, Baines 2008, and Boyce & DiPrima 2009).

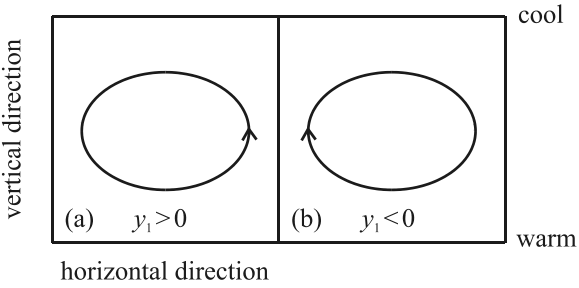



**Fig. 9.8.** An illustration of convection according to Lorenz's equations. Fluid flow is considered in a single cell (i.e., in one box). The flow is heated from below and cooled from above, and there are slippery nonconducting side walls. For a sufficiently large temperature difference the warmer fluid rises, cools down at the top, and moves downwards. This results in a steady convective motion. The strength and direction of this circulation is measured by the variable $y_1$.

Figure 9.8 shows an illustration of the case considered. The model variables have the following meaning. The strength and direction of the circulation is measured by $y_1$, $y_2$ measures the horizontal temperature variation, and $y_3$ measures the vertical temperature variation. The model parameter $Pr$ is the Prandtl number (the ratio of diffusivities of momentum and heat), and $b$ is defined by $b = 4/(1 + a^2)$. The parameter $a$ is a horizontal wavenumber for the convection cells, and $b$ measures the width-to-height ratio of the convection layer. The most relevant model parameter is $R$, which is proportional to the vertical temperature difference (the driving force of the system). In particular, $R$ is defined by $R = R_a / R_c$, where $R_a$ refers to the Rayleigh number and $R_c$ refers to the critical value of the Rayleigh number (the Rayleigh number that is required for the onset of convection). A large value of $R$ implies a large thermal forcing of motion.

**Lorenz Model Considered.** To simplify the relatively complicated analysis of Lorenz's equations (9.144) we will follow the studies of Saltzman and Lorenz by specifying $a^2 = 1/2$ so that $b = 8/3$. We also specify $Pr = 10$, which is a realistic value for water. The equation system that results from these assumptions reads

$$\frac{dy_1}{dt} = 10(y_2 - y_1), \tag{9.145a}$$

$$\frac{dy_2}{dt} = y_1(R - y_3) - y_2, \tag{9.145b}$$

$$\frac{dy_3}{dt} = y_1 y_2 - \frac{8}{3} y_3. \tag{9.145c}$$

These equations will be analyzed in the following in dependence on $R \geq 0$, which controls the amount of thermal forcing.

### 9.5.2 Linear Stability Analysis

**Equilibrium Points.** Linear stability analysis has to be applied to derive analytical conclusions regarding the nonlinear equation system (9.145). Such analysis requires the calculation of equilibrium points. These points are defined by

$$0 = 10(y_2 - y_1), \tag{9.146a}$$

$$0 = y_1(R - y_3) - y_2, \tag{9.146b}$$

$$0 = y_1 y_2 - \frac{8}{3} y_3. \tag{9.146c}$$

Equation (9.146a) implies $y_2 = y_1$. Hence, the other two conditions can be written

$$0 = y_1(R - y_3 - 1), \tag{9.147a}$$

$$0 = y_1^{\,2} - \frac{8}{3} y_3. \tag{9.147b}$$

The first way to satisfy Eq. (9.147a) is given by $y_1 = 0$, which implies $y_2 = y_3 = 0$. The second possibility to satisfy Eq. (9.147a) is $y_3 = R - 1$. This setting implies that $y_2 = y_1 = [8\,(R - 1)/3]^{1/2}$, or $y_2 = y_1 = -[8\,(R - 1)/3]^{1/2}$. Hence, we have three equilibrium points given by

$$\begin{aligned} &P_1 = (0,0,0), \qquad P_2 = \left(\sqrt{8(R-1)/3}, \sqrt{8(R-1)/3}, R-1\right), \\ &P_3 = \left(-\sqrt{8(R-1)/3}, -\sqrt{8(R-1)/3}, R-1\right). \end{aligned} \tag{9.148}$$

The properties of these equilibrium points depend on $R$. For $R < 1$, the only real equilibrium point is given by $P_1$. For $R > 1$, there are three real equilibrium points. For $R = 1$ we have three times the equilibrium point (0, 0, 0).

**Linear Stability Analysis.** The linear equation system in the neighborhood of any equilibrium point can be obtained by generalizing the equation system (9.69) to the three-dimensional case,

$$\frac{d}{dt}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \\ y_3 - Y_3 \end{pmatrix} = \begin{pmatrix} \partial F_1/\partial y_1 & \partial F_1/\partial y_2 & \partial F_1/\partial y_3 \\ \partial F_2/\partial y_1 & \partial F_2/\partial y_2 & \partial F_2/\partial y_3 \\ \partial F_3/\partial y_1 & \partial F_3/\partial y_2 & \partial F_3/\partial y_3 \end{pmatrix}_{\substack{y_1=Y_1 \\ y_2=Y_2 \\ y_3=Y_3}} \begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \\ y_3 - Y_3 \end{pmatrix}. \tag{9.149}$$

Here, $F_1$, $F_2$, and $F_3$ represent the right-hand sides of the three equations (9.145), respectively. Regarding Eqs. (9.145) considered we find

$$\frac{d}{dt}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \\ y_3 - Y_3 \end{pmatrix} = \begin{pmatrix} -10 & 10 & 0 \\ R - Y_3 & -1 & -Y_1 \\ Y_2 & Y_1 & -8/3 \end{pmatrix} \begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \\ y_3 - Y_3 \end{pmatrix}. \tag{9.150}$$

**First Equilibrium Point.** The linear equation system for the dynamics near the first equilibrium point $P_1 = (Y_1, Y_2, Y_3) = (0, 0, 0)$ is then given by

$$\frac{d}{dt}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \\ y_3 - Y_3 \end{pmatrix} = \begin{pmatrix} -10 & 10 & 0 \\ R & -1 & 0 \\ 0 & 0 & -8/3 \end{pmatrix}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \\ y_3 - Y_3 \end{pmatrix}. \tag{9.151}$$

The eigenvalues of this equation system follow from the extension of Eq. (9.21) to the three-dimensional case considered,

$$\begin{aligned} 0 &= \begin{vmatrix} -10-r & 10 & 0 \\ R & -1-r & 0 \\ 0 & 0 & -8/3-r \end{vmatrix} = -(10+r)(1+r)(8/3+r) + 10R(8/3+r) \\ &= -(8/3+r)\left[(10+r)(1+r) - 10R\right] = -(8/3+r)\left[r^2 + 11r - 10(R-1)\right]. \end{aligned} \tag{9.152}$$

This cubic equation for $r$ has three roots that are determined by $8/3 + r = 0$ and the condition that the bracket term is equal to zero,

$$\begin{aligned} r_1 &= -\frac{8}{3}, \\ r_2 &= -\frac{11}{2} + \sqrt{\frac{121}{4} + 10(R-1)} = -\frac{11}{2}\left(1 - \sqrt{1 + \frac{40}{121}(R-1)}\right), \\ r_3 &= -\frac{11}{2} - \sqrt{\frac{121}{4} + 10(R-1)} = -\frac{11}{2}\left(1 + \sqrt{1 + \frac{40}{121}(R-1)}\right). \end{aligned} \tag{9.153}$$

The eigenvalues $r_1$ and $r_3$ are always negative. The sign of $r_2$ does depend on $R$: we have $r_2 < 0$ if $R < 1$, and $r_2 > 0$ if $R > 1$.

**Second and Third Equilibrium Points. I.** We can combine the stability analysis for the second and third equilibrium point by writing Eq. (9.150) as

$$\frac{d}{dt}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \\ y_3 - Y_3 \end{pmatrix} = \begin{pmatrix} -10 & 10 & 0 \\ 1 & -1 & -s\sqrt{8(R-1)/3} \\ s\sqrt{8(R-1)/3} & s\sqrt{8(R-1)/3} & -8/3 \end{pmatrix}\begin{pmatrix} y_1 - Y_1 \\ y_2 - Y_2 \\ y_3 - Y_3 \end{pmatrix}. \tag{9.154}$$

The settings $s = 1$ and $s = -1$ correspond to the consideration of $P_2$ and $P_3$, respectively. The eigenvalues of this system are determined by the condition

$$0 = \begin{vmatrix} -10-r & 10 & 0 \\ 1 & -1-r & -s\sqrt{8(R-1)/3} \\ s\sqrt{8(R-1)/3} & s\sqrt{8(R-1)/3} & -8/3-r \end{vmatrix}. \tag{9.155}$$

**Table 9.2** Solutions $y = r + a/3$ of the reduced cubic equation (9.159). The case $\alpha > 0$ implies $D > 0$. The last row provides the formula for the calculation of $\varphi$.

| | $\alpha < 0$: $D \le 0$ | $\alpha < 0$: $D > 0$ | $\alpha > 0$: $D > 0$ |
|---|---|---|---|
| $y_1$ | $-2E\cos\frac{\varphi}{3}$ | $-2E\cosh\frac{\varphi}{3}$ | $-2E\sinh\frac{\varphi}{3}$ |
| $y_2$ | $-2E\cos\frac{\varphi+2\pi}{3}$ | $E\left[\cosh\frac{\varphi}{3}+i\sqrt{3}\sinh\frac{\varphi}{3}\right]$ | $E\left[\sinh\frac{\varphi}{3}+i\sqrt{3}\cosh\frac{\varphi}{3}\right]$ |
| $y_3$ | $-2E\cos\frac{\varphi+4\pi}{3}$ | $E\left[\cosh\frac{\varphi}{3}-i\sqrt{3}\sinh\frac{\varphi}{3}\right]$ | $E\left[\sinh\frac{\varphi}{3}-i\sqrt{3}\cosh\frac{\varphi}{3}\right]$ |
| $\varphi$ | $\cos\varphi=\frac{\beta}{2E^3}$ | $\cosh\varphi=\frac{\beta}{2E^3}$ | $\sinh\varphi=\frac{\beta}{2E^3}$ |

The expansion of the determinant provides the condition

$$\begin{aligned}
0 &= -(10+r)\left[(1+r)(8/3+r)+8(R-1)/3\right]+10\left[8/3+r-8(R-1)/3\right] \\
&= -(8/3+r)\left[(10+r)(1+r)-10\right]-8(R-1)/3\left[10+10+r\right] \\
&= -r(8/3+r)\left[r+11\right]-8(R-1)/3\left[20+r\right] \\
&= -r\left(r^2+\frac{8+33}{3}r+11\frac{8}{3}+\frac{8(R-1)}{3}\right)-\frac{160(R-1)}{3} \\
&= r^3+\frac{41}{3}r^2+\frac{8(R+10)}{3}r+\frac{160(R-1)}{3}.
\end{aligned} \tag{9.156}$$

This eigenvalue equation is independent of $s$. Hence, this equation is the same for the equilibrium points $P_2$ and $P_3$.

**Solutions of Cubic Equations.** Before analyzing the consequences of the last equation, let us briefly review the solutions of cubic equations. We consider the equation

$$0 = r^3 + ar^2 + br + c. \tag{9.157}$$

Here, $a$, $b$, and $c$ are any real coefficients. By introducing $y = r + a/3$, Eq. (9.157) can be written as a reduced equation that does not contain a quadratic term,

$$\begin{aligned}
0 &= \left(y-\frac{a}{3}\right)^3+a\left(y-\frac{a}{3}\right)^2+b\left(y-\frac{a}{3}\right)+c \\
&= y^3-3y^2\frac{a}{3}+3y\left(\frac{a}{3}\right)^2-\left(\frac{a}{3}\right)^3+ay^2-2ay\frac{a}{3}+a\left(\frac{a}{3}\right)^2+by-b\frac{a}{3}+c \\
&= y^3+y\left[3\left(\frac{a}{3}\right)^2-6\left(\frac{a}{3}\right)^2+b\right]-\left(\frac{a}{3}\right)^3+3\left(\frac{a}{3}\right)^3-b\frac{a}{3}+c.
\end{aligned} \tag{9.158}$$

A more convenient way is to write this equation as

$$0 = y^3 + \alpha y + \beta, \tag{9.159}$$

where $\alpha$ and $\beta$ are defined by

$$\alpha = b - 3\left(\frac{a}{3}\right)^2, \qquad \beta = 2\left(\frac{a}{3}\right)^3 - b\frac{a}{3} + c. \tag{9.160}$$

The three solutions of Eq. (9.159) are given in Table 9.2 in dependence on

$$D = \left(\frac{\alpha}{3}\right)^3 + \left(\frac{\beta}{2}\right)^2, \qquad E = \frac{\beta}{|\beta|}\sqrt{\frac{|\alpha|}{3}}. \tag{9.161}$$

Here, $D$ is called the discriminant. The solutions of the original Eq. (9.157) can be obtained by means of the relation $r = y - a/3$. Table 9.2 shows that there are two possibilities: there can be either three real solutions or one real solution and two conjugate complex solutions. The two solution regimes are separated at $D = 0$: for $D \le 0$ we have three real solutions, and for $D > 0$ we have one real and two complex solutions.

**Pure Imaginary Eigenvalues.** The two complex solutions may have positive or negative real parts. Regarding the evaluation of stability it is relevant to know for which $R$ the real parts of complex roots are equal to zero (because we know then for which $R$ the real parts of complex roots are positive and negative). This specific case of pure imaginary roots is given under the conditions that $c = a\,b$ and $b > 0$. To prove the requirement of the first condition $c = a\,b$ we write Eq. (9.157) for this case as

$$0 = r^3 + a r^2 + b r + a b = (r + a)(r^2 + b). \tag{9.162}$$

This representation reveals the roots $r_1 = -a$ and $r_{2,3} = \pm\, i\, b^{1/2}$. The comparison of the solutions $r_1 = -a$ and $r_{2,3} = \pm\, i\, b^{1/2}$ for this case with the solutions presented in Table 9.2 shows that $E\cosh(\varphi/3)$ or $E\sinh(\varphi/3)$, which represent the real parts of $y_2$ and $y_3$ depending on a negative or positive sign of $\alpha$, respectively, must be equal to $a/3$. These values imply zero real parts of $r_2$ and $r_3$ according to $r = y - a/3$. The second condition $b > 0$ is a requirement to have a discriminant $D > 0$, which is needed for the existence of complex solutions. We have to calculate $D$ to show the correctness of this claim. For $c = a\,b$, $\alpha$ and $\beta$ are given by

$$\alpha = b - 3\left(\frac{a}{3}\right)^2, \qquad \beta = 2\left(\frac{a}{3}\right)^3 + 2b\frac{a}{3}. \tag{9.163}$$

Hence, the discriminant $D$ is given by

$$D = \left(\frac{\alpha}{3}\right)^3 + \left(\frac{\beta}{2}\right)^2 = \left(\frac{b}{3} - \left(\frac{a}{3}\right)^2\right)^3 + \left(\left(\frac{a}{3}\right)^3 + b\frac{a}{3}\right)^2. \tag{9.164}$$

The evaluation of this expression provides

$$D = \left(\frac{b}{3}\right)^3 - 3\left(\frac{b}{3}\right)^2\left(\frac{a}{3}\right)^2 + 3\frac{b}{3}\left(\frac{a}{3}\right)^4 - \left(\frac{a}{3}\right)^6 + \left(\frac{a}{3}\right)^6 + 2b\left(\frac{a}{3}\right)^4 + b^2\left(\frac{a}{3}\right)^2$$
$$= \left(\frac{b}{3}\right)^3 + 6\left(\frac{b}{3}\right)^2\left(\frac{a}{3}\right)^2 + 9\frac{b}{3}\left(\frac{a}{3}\right)^4 = \frac{b}{3}\left(\frac{b}{3} + 3\left(\frac{a}{3}\right)^2\right)^2. \tag{9.165}$$

Therefore, $D > 0$ under the condition that $b > 0$.

**Second and Third Equilibrium Points. II.** The eigenvalue equation (9.156) can be analyzed on the basis of the solutions of cubic equations described in the preceding two paragraphs. Let us prepare this discussion by the calculation of two characteristic values of $R$. A first characteristic value $R_1$ is the value that separates three real solutions from one real solution and two conjugate complex solutions. $R_1$ is determined by $D = (\alpha/3)^3 + (\beta/2)^2 = 0$. For our case, $\alpha$ and $\beta$ are given by

$$\alpha = \frac{8(R+10)}{3} - 3\left(\frac{41}{9}\right)^2 = \frac{8}{3}R - \frac{961}{27},$$
$$\beta = 2\left(\frac{41}{9}\right)^3 - \frac{8(R+10)}{3}\frac{41}{9} + \frac{160(R-1)}{3} = \frac{1112}{27}R + \frac{10402}{729}. \tag{9.166}$$

The use of the latter relations in $D = (\alpha/3)^3 + (\beta/2)^2$ provides the expression

$$D(R) = \left(\frac{8}{9}R - \frac{961}{81}\right)^3 + \left(\frac{556}{27}R + \frac{5201}{729}\right)^2 = 32\left(R - \frac{961}{72}\right)^3 + \left(139R + \frac{5201}{108}\right)^2$$
$$= R^3 - 3R^2\frac{961}{72} + 3R\frac{961^2}{72^2} - \frac{961^3}{72^3} + \frac{139^2}{32}R^2 + 139R\frac{5201}{16\cdot 108} + \frac{5201^2}{32\cdot 108^2}$$
$$= R^3 + 2\frac{54{,}119}{6\cdot 4\cdot 8}R^2 + 108\frac{15{,}245}{16\cdot 108}R - 2\cdot 108^2\frac{36{,}885}{32\cdot 108^2}$$
$$= R^3 + \frac{54{,}119}{96}R^2 + \frac{15{,}245}{16}R - \frac{36{,}885}{16}. \tag{9.167}$$

The solution of cubic equations described above shows that the equation $D(R) = 0$ has two negative real roots and one positive real root. The negative roots can be disregarded because we consider $R \geq 0$. The real root, which is $R_1$, is given by

$$R_1 = -2\frac{\beta}{|\beta|}\sqrt{\frac{|\alpha|}{3}}\cos\left(\frac{1}{3}\arccos\left(\frac{|\beta|3^{3/2}}{2|\alpha|^{3/2}}\right) + \frac{2\pi}{3}\right) - \frac{a}{3} \approx 1.3456. \tag{9.168}$$

Here, $\alpha = b - 3(a/3)^2$ and $\beta = 2(a/3)^3 - ab/3 + c$, where $a$, $b$, and $c$ are the coefficients of $R^2$, $R$, and the last term in the cubic equation (9.156), respectively. The validity of this value may be seen by proving that $R_1 = 1.3456$ implies $D = 0$.

**Table 9.3** The linear stability properties of the Lorenz equations (9.145).

| | $P_1$ | $P_2$ and $P_3$ |
|---|---|---|
| $0 < R < 1$ | asymptotically stable:<br>3 real negative roots | do not exist |
| $1 < R < R_1$ | unstable:<br>2 real negative roots,<br>1 real positive root | asymptotically stable:<br>3 real negative roots |
| $R_1 < R < R_2$ | unstable:<br>2 real negative roots,<br>1 real positive root | asymptotically stable:<br>1 real negative root,<br>2 complex roots with negative real parts |
| $R_2 < R$ | unstable:<br>2 real negative roots,<br>1 real positive root | unstable:<br>1 real negative root,<br>2 complex roots with positive real parts |

A second characteristic value $R_2$ of $R$ is the value that separates complex roots with negative and positive real parts. In terms of the notation of the general cubic equation (9.157), this value is determined by the condition $c = a\,b$. By adopting $a$, $b$, and $c$ according to the eigenvalue equation (9.156), the condition for $R_2$ reads

$$\frac{41}{3}\frac{8(R_2+10)}{3} = \frac{160(R_2-1)}{3}. \tag{9.169}$$

This equation is solved by

$$R_2 = \frac{470}{19} \approx 24.7368. \tag{9.170}$$

**Second and Third Equilibrium Points. III.** The stability behavior near $P_2$ and $P_3$ can be determined now by taking reference to $R_1$ and $R_2$. We find the following features (see also the summary of linear stability features in Table 9.3):

a) $1 < R < R_1$: All the terms in Eq. (9.156) are positive for $R > 1$. Hence, all real solutions have to be negative. The discriminant $D = (\alpha/3)^3 + (\beta/2)^2$ increases according to Eq. (9.167), $dD/dR > 0$, and we know that $D = 0$ at $R_1$. Thus, $D < 0$ for $1 < R < R_1$, which means that there are three real negative roots for $1 < R < R_1$. Hence, the solution near $P_2$ and $P_3$ is asymptotically stable.

b) $R_1 < R < R_2$: For $R_1 < R$ we have one real root, which has to be negative because all real roots must be negative, and two conjugate complex roots. The complex roots arise from imaginary contributions that appear in addition to the negative real parts of these roots. The real parts of these roots are negative when $R < R_2$. Hence, the solution near $P_2$ and $P_3$ is asymptotically stable in this regime, too.

c) $R_2 < R$: For $R$ values in this regime we have one negative real root and two conjugate complex roots. The complex roots do have now positive real parts because of $R_2 < R$. Hence, the solution near $P_2$ and $P_3$ is unstable in this regime.

### 9.5.3 Deterministic Chaos

**Equations Considered.** The illustration of characteristic features of the Lorenz equations (9.145) will be focused on the case $R_2 < R$. A discussion of solution properties for other cases can be found, e.g., in Sparrow (1982). In particular, we use $R = 28$, which means we consider the equations

$$\frac{dy_1}{dt} = 10(y_2 - y_1), \tag{9.171a}$$

$$\frac{dy_2}{dt} = y_1(28 - y_3) - y_2, \tag{9.171b}$$

$$\frac{dy_3}{dt} = y_1 y_2 - \frac{8}{3} y_3. \tag{9.171c}$$

The reason for the consideration of these equations is that this case, for which we do not have asymptotically stable solutions, is the most interesting one. It will be shown below that the solutions exhibit a chaotic behavior in this case. This model may be considered as a highly simplified model for turbulence that is described by the Navier-Stokes equations. The condition $R_2 < R$ does appear here in analogy to the condition for the onset of turbulence that the Reynolds number must be above a critical Reynolds number.

**Numerical Solution.** Due to the nonlinear terms involved, the nonlinear equation system (9.171) can only be solved numerically. For doing this we write these equations as a system of difference equations

$$y_1^{n+1} = y_1^n + 10\,\Delta t\left(y_2^n - y_1^n\right), \tag{9.172a}$$

$$y_2^{n+1} = y_2^n + \Delta t\left(y_1^n(28 - y_3^n) - y_2^n\right), \tag{9.172b}$$

$$y_3^{n+1} = y_3^n + \Delta t\left(y_1^n y_2^n - \frac{8}{3} y_3^n\right), \tag{9.172c}$$

where $n = 0, 1, 2, \ldots$ Starting from the initial values $(y_1^0, y_2^0, y_3^0) = (y_{10}, y_{20}, y_{30})$, these equations describe the evolution of $y_1, y_2, y_3$ in time $t = n\,\Delta t$. The initial data $(y_{10}, y_{20}, y_{30}) = (5, 5, 5)$ will be applied here (except for the study of the influence of varying initial data described below). Equations (9.172) introduces a parameter: the time interval $\Delta t$. This time interval is considered to be sufficiently small in order to produce solutions of Eqs. (9.172) that are independent of $\Delta t$. In that case, the solutions of Eqs. (9.172) are seen as solutions of the differential equation system (9.171). The effect of different $\Delta t$ settings is illustrated in Fig. 9.9. This figure shows that the solutions of Eqs. (9.172) do not become independent of $\Delta t$ for $\Delta t$ variations over seven orders of magnitude ($\Delta t = 10^{-3}$ to $\Delta t = 10^{-10}$). It is not easy to consider the effect of smaller $\Delta t$ values because such simulations are expensive.

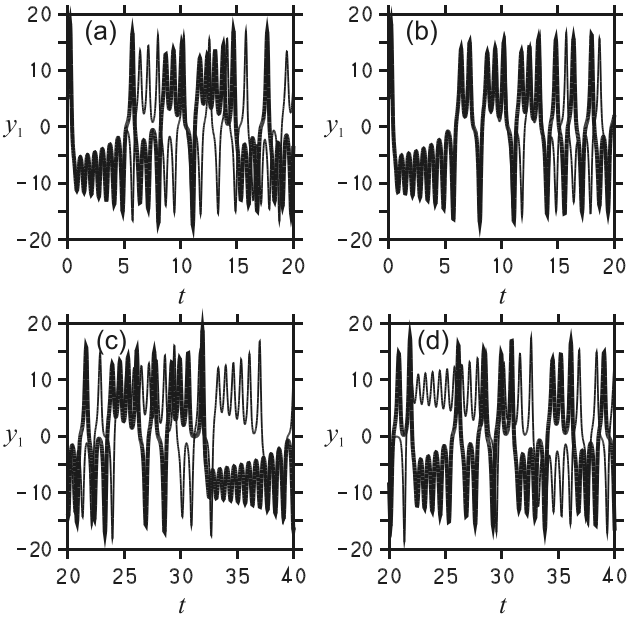



**Fig. 9.9.** Solutions $y_1(t)$ of the Lorenz equations (9.172) in dependence on the time interval $\Delta t$, where $(y_{10}, y_{20}, y_{30}) = (5, 5, 5)$. The *thick lines* present the results for $\Delta t = 10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}$ in (a)–(d), respectively; the *thin lines* show the results for $\Delta t = 10^{-4}, 10^{-6}, 10^{-8}, 10^{-10}$ in (a) –(d), respectively.

The simulation, e.g., for $\Delta t = 10^{-10}$ requires 11.1 hours on a Pentium(R) 4 CPU 3.2 GHz personal computer with 1 GB memory. To improve the clarity of these plots, $y_1$ is only shown from $t = 20$ to $t = 40$ in Figs. 9.9c-d. For $t < 20$, the thick and thin lines in these figures do hardly show any difference. The conclusion that the solutions of Eqs. (9.172) depend on the time interval $\Delta t$ is surprising because this observation differs from the behavior of many other differential equations. It is relevant to note that this conclusion does not depend on the simple numerical scheme (9.172) used to solve the Lorenz equations (9.171): the result is the same for a variety of more advanced numerical schemes (see Yao 2007, 2010, Yao & Hughes 2008, and Liao 2009). Hence, it may be impossible to obtain a unique solution of the Lorenz equations (at least, no such solution that is independent of $\Delta t$ has been reported so far). The Lorenz equations (9.145) have to be considered, therefore, as a guideline for the construction of numerical schemes (one possible numerical scheme is given by Eqs. (9.172)) that are defined in conjunction with a specific choice of $\Delta t$. This feature of the Lorenz equations also poses questions about the reliability of numerical solutions of the fluid dynamics equations (the Navier-Stokes equations): see Yao (2007, 2010).

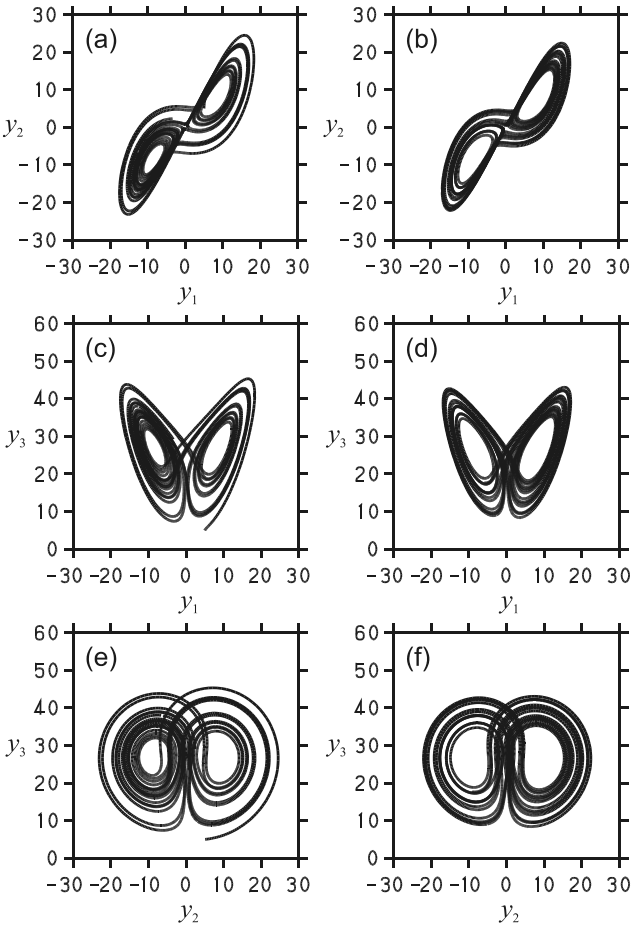



**Fig. 9.10.** The evolution of variables determined by the Lorenz equations (9.172) in several phase planes, where $\Delta t = 10^{-3}$ and $(y_{10}, y_{20}, y_{30}) = (5, 5, 5)$. (a), (c), and (e) show the evolution from $t = 0$ to $t = 20$; (b), (d), and (f) show the evolution from $t = 10{,}000$ to $t = 10{,}020$.

**Phase Plane Evolution.** The phase plane evolution of $y_1$, $y_2$, and $y_3$ is shown in Fig. 9.10 for $\Delta t = 10^{-3}$, which will be used in the following. These trajectories do never cross each other because the system never exactly repeats itself. There is no convergence to any asymptotically stable state: the trajectories from $t = 10{,}000$ to $t = 10{,}020$ are very similar to the trajectories from $t = 0$ to $t = 20$. The long-term behavior of the phase plane trajectories shown in these plots is called the Lorenz attractor, noted for the butterfly shape of the $y_1$-$y_3$ trajectory and the owl mask shape of the $y_2$-$y_3$ trajectory.

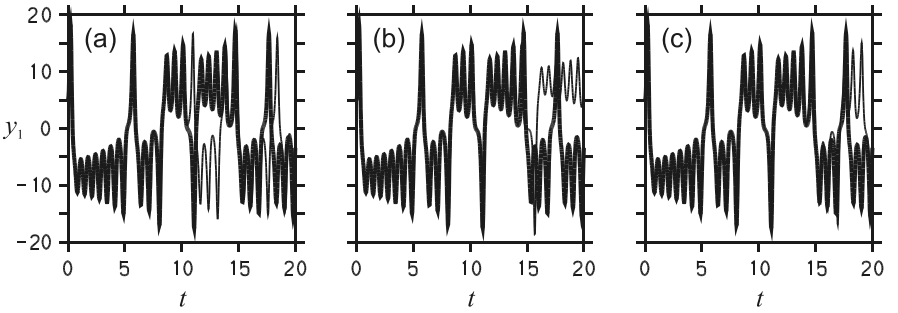



**Fig. 9.11.** Solutions $y_1(t)$ of the Lorenz equations (9.172): the influence of variations of initial conditions, where $\Delta t = 10^{-3}$. The *thick lines* in these figures show the result for $(y_{10}, y_{20}, y_{30}) = (5, 5, 5)$. The *thin lines* in (a), (b), and (c) result from the initial data (5.01, 5, 5), (5.001, 5, 5), and (5.0001, 5, 5), respectively.

**The Butterfly Effect.** The Lorenz equations are not only sensitive to variations of the time step used to integrate these equations. These equations are also very sensitive to minor variations of the initial conditions, as demonstrated in Fig. 9.11. In fact, every difference in initial data will result in different solutions. Figure 9.11 demonstrates that the larger the difference of initial data, the sooner there will be a difference between solutions. Lorenz did accidentally discover this sensi-tivity of solutions to perturbations of initial data when he restarted the numerical integration of equations by rounding-off the data values used in the computations. The sensitive dependence of solutions on initial conditions is called the Butterfly Effect because Lorenz compared this dependence with the effect of a butterfly on the weather: he asked "Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?". The Lorenz equations are clearly a highly simplified version of equations that can be used for weather forecasting. However, they may explain the reason of why long-term weather predictions are simply impossible: the nonlinear interactions of variables involved in such equations imply a high sensitivity to perturbations, and perturbations have to be always taken into account (initial data are never known exactly).

**Probability Density Functions.** The Lorenz equations are deterministic, but they produce output that looks like random data. It is, therefore, a reasonable idea to study the probability for finding certain solution values. This question will be addressed by considering the probability density function (PDF) of $y_1$, $y_2$, and $y_3$ values. The latter PDFs are denoted by $f_1(x_1)$, $f_2(x_2)$, and $f_3(x_3)$, where $x_1$, $x_2$, and $x_3$ represent the sample space variables of $y_1$, $y_2$, and $y_3$, respectively. The Lorenz equations were solved with $\Delta t = 10^{-3}$ up to $t = 25$, $t = 100$, and $t = 1000$. For $t \leq 15$ the PDFs are still heavily affected by randomness: there are very sharp peaks that are difficult to resolve. The state at $t = 25$ corresponds to a state at which the PDF

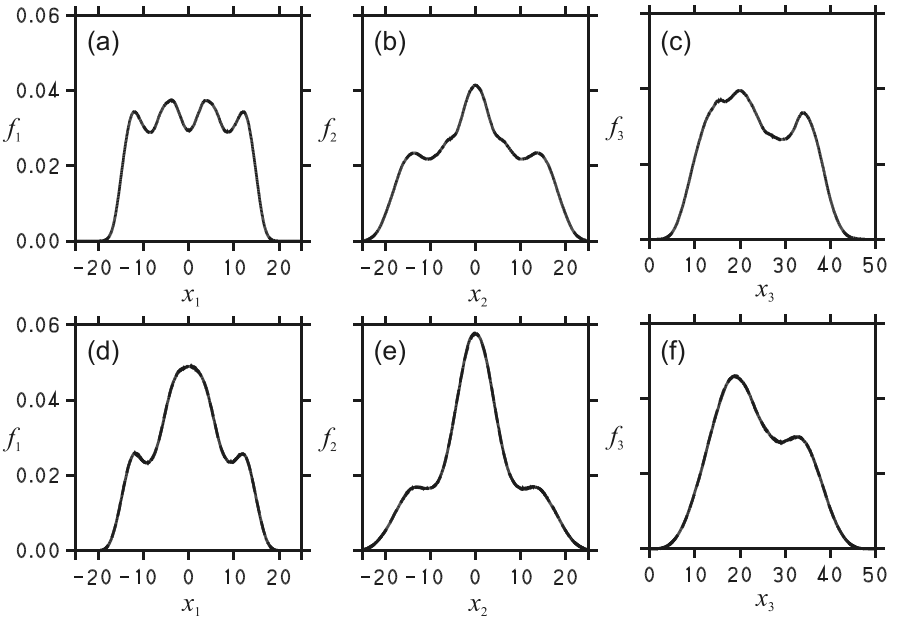



**Fig. 9.12.** The probability density function (PDF) $f_1(x_1)$, $f_2(x_2)$, and $f_3(x_3)$ related to $y_1$, $y_2$, and $y_3$ values, respectively. The *upper row* shows these PDFs at $t = 25$, and the *lower row* shows the PDFs at $t = 100$ (*dashed line*) and $t = 1000$ (*solid line*).

is relatively smooth but still significantly affected by the initial data. The state at $t = 100$ represents the asymptotic state. Evidence for this conclusion is provided by the PDF at $t = 1000$ which does not show an observable differences to the PDF at $t = 100$. $10^6$ solutions were generated by using varying initial conditions for $y_1$ (with an equal distance) between 4.995 and 5.005. The PDFs were calculated as filtered PDFs according to the explanations in Chap. 4. A filter interval equal to 2 was used to obtain smooth PDF curves. The results are shown in Fig. 9.12 for $t = 25$, $t = 100$, and $t = 1000$. The PDFs at $t = 25$ are characterized by several modes (4, 5, and 3 modes regarding the $f_1$, $f_2$, and $f_3$ curves). These PDF structures show that the initial data considered (which do only involve variations of $y_1$ values) may excite a spectrum of different motions. At $t = 100$ (i.e., in the asymptotic stage) we observe a smoothening between these modes: different modes merge. The latter leads to the development of a central mode, but the outer modes are still present. The PDF curves seen here are clearly different from the features of velocity and temperature PDFs for the unstably stratified atmospheric boundary layer (only the $f_3$ PDF shows some similarities at $t = 100$), but the $f_1$, $f_2$, and $f_3$ curves are similar in the sense that they represent a superposition of several distinct motions.

## 9.6 Summary

Let us summarize the observations made in this chapter regarding the extension of laws for one variable to the multivariate case of several interacting variables. This will be done by addressing the questions posed at the end of Sect. 9.1, i.e., the questions about the formulation of laws and the use of such equations.

**Formulation of Multivariate Laws.** Newton's Laws of Mechanics apply to the case of several variables. Thus, there is no question about the laws for mechanical processes of macroscopic bodies that move with velocities much slower than the speed of light. With regard to the laws for population dynamics we have another case because there is no unique formulation of such processes (see the discussions in Chap. 7). In this case, we follow the spirit of formulating such laws by using empirical modifications of single-variable equations. The particular question with regard to both the laws for population ecology and mechanical processes is how it is possible to use such equations for several variables. This problem is much more complicated than the use of equations for only one variable. Let us summarize the findings obtained regarding this question.

**Numerical Solution of Multivariate Equations.** The numerical solution of equations represents a general methodology for using evolution equations in order to study the features of processes. This approach does work if it is possible to find convergent solutions. A convergent solution represents a solution that is independent of variations of small time intervals used in the numerical scheme. For most equations it is possible to find such convergent solutions, but this is not always the case. An example for the latter case was given here by the Lorenz equations. So far, a convergent solution has not been reported for these equations. This finding does have implications for practical problems. The Lorenz equations represent a highly simplified version of the Navier-Stokes equations that are used to calculate fluid dynamics processes. With regard to most applications it is very expensive to prove the convergence of solutions to the Navier-Stokes equations (such simulations may need several years). Thus, solutions of the Navier-Stokes equations are calculated by adopting a relatively small time step. Then, there is the question of whether such solutions represent convergent solutions, which is not the case for the simple Lorenz model derived from the Navier-Stokes equations.

**Analytical Study of Multivariate Equations.** It is hardly possible to integrate multivariate nonlinear coupled equations. Usually, analytical conclusions can only be derived by means of linear stability analysis. Such analyses are very helpful, as demonstrated for the examples given in this chapter. We obtain insight in this way that can hardly be obtained by numerical simulations. It would be scarcely possible, e.g., to use numerical solutions for accurate calculations of the critical number $R_2$ that separates nonchaotic and chaotic solutions of the Lorenz equations. On the

other hand, linear stability theory does not represent an alternative to numerical solutions, because the overall features of nonlinear equations cannot be studied in this way. The analysis of nonlinear equations is only possible under very special conditions. An example was given here by the discussion of the application of Lyapunov's second method. Such methods are applicable if there is a way to find conserved variables, as, e.g., the total energy of a process.

## 9.7 Exercises

**9.2.1** Consider the linear equation system

$$\frac{d}{dt}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} -1 & 2 \\ 4 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

The initial values are given by $y_1(0) = y_{10}$ and $y_2(0) = y_{20}$, where $y_{10}$ and $y_{20}$ are any parameters.

a) Determine the solutions $y_1(t)$ and $y_2(t)$ in dependence on $y_{10}$ and $y_{20}$.

b) Which relation between the initial values $y_{10}$ and $y_{20}$ is required such that $y_2$ is a linear function of $y_1$ that disappears asymptotically? Find the corresponding linear function $y_2 = y_2(y_1)$.

c) Which relation between the initial values $y_{10}$ and $y_{20}$ is required such that $y_2$ is a linear function of $y_1$ that goes to infinity asymptotically? Find the corresponding linear function $y_2 = y_2(y_1)$.

**9.2.2** Consider the linear equation system

$$\frac{d}{dt}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & -A(1+A) \\ -2 & A \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

The initial values are $y_1(0) = 1$ and $y_2(0) = -1$, and $A$ is any parameter.

a) Find the solutions $y_1(t)$ and $y_2(t)$ to this initial value problem.

b) For which range of $A$ values do the solutions become zero as $t \to \infty$?

**9.2.3** Consider the nonlinear equation system

$$\frac{d}{dt}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & -y_2 \\ 1 & -y_1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

a) Find the equilibrium points of this equation system.

b) Which of the equilibrium points will be realized? Apply linear stability analysis to address this question.

**9.2.4** Consider the nonlinear equation system

$$\frac{d}{dt}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_1 & y_2 \\ y_2 & y_1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

a) Find the equilibrium points of this equation system.
b) Which of the equilibrium points will be realized? Apply linear stability analysis to address this question.
c) Apply the solutions $y_1(t)$ and $y_2(t)$ to the nonlinear equation system to address again the question about the realization of equilibrium solutions. Hint: you may use the equation system to derive and solve equations for $y_1 + y_2$ and $y_1 - y_2$, respectively.

**9.2.5** A specific form of Duffing's nonlinear spring model reads (Wiggins 2010)

$$\frac{d}{dt}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 - y_1^{\,2} & 0 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

a) Find the equilibrium points of this equation system.
b) Show that two equilibrium points represent centers.
c) Use the linear equation systems near the centers to find the shape of trajectories in the $y_1$-$y_2$ phase plane. Hint: you may calculate the ratio $dy_1/dy_2 = (dy_1/dt)/(dy_2/dt)$ and solve the resulting separable equation.
d) Explain the type of equation obtained for trajectories in c).

**9.3.1** Consider the following modification of the competition for food dynamics (9.73) discussed in Sect. 9.3.2 ($d$ is a non-negative number),

$$\frac{dy_1}{dt} = y_1\left(1 - y_1 - d\,y_2\right), \qquad \frac{dy_2}{dt} = y_2\left(1 - d\,y_1\right).$$

a) Find the equilibrium points of this equation system.
b) Determine the stability behavior of solutions near the equilibrium points in dependence on the model parameter $d$.
c) The coexistence equilibrium point depends on the value of $d$. For which range of $d$ do we find a non-negative coexistence equilibrium point?
d) Consider the range of values of $d$ determined in c). Which equilibrium point will be realized asymptotically?

**9.3.2** Consider the following modified Lotka-Volterra equations for the prey $y_1$ (food fish) and predators $y_2$ (sharks): see Allen (2007),

$$\frac{dy_1}{dt} = y_1\left(A - f - B\,y_2\right), \qquad \frac{dy_2}{dt} = y_2\left(-C - f + D\,y_1\right).$$

Here, $A$, $B$, $C$, $D$, and $f$ are non-negative constants. The parameter $f$ models a prey reduction due to fishing. How does $f$ affect a coexistence of species?

**9.3.3** Infectious diseases such as measles, mumps, rubella, and chickenpox are modeled by involving three groups of individuals (Kermack & McKendrick 1927, Anderson & May 1979a, 1979b, Anderson 1982, Fulford et al. 1997, Edelstein-Keshet 2005, Allen 2007). The total population $N$, which is considered to be constant, is subdivided into susceptible ($S$), infective ($I$), and removed ($R$) classes: $N = S(t) + I(t) + R(t)$. Susceptible refers to individuals not infected but who are capable of contracting the disease and becoming infective. Infective refers to individuals who are infected and infectious. Removed refers to individuals who have had the disease and have definitely recovered, who are permanently immune, or are isolated until recovery. A very simple epidemic model (the $SI$ model) assume $R = 0$ and relates $S$ and $I$ by

$$\frac{dS}{dt} = -\frac{\beta}{N} S I, \qquad \frac{dI}{dt} = \frac{\beta}{N} S I.$$

Here, $\beta$ is a positive constant of proportionality.

a) Use the relation $N = S + I$ to derive a closed equation for $I$. Compare this equation with differential equations considered in Chap. 7. Which type of equation is the equation for $I$?

b) Solve the differential equation for $I$.

c) Calculate the asymptotic values of $S$ and $I$ for large values of $t$. Explain the meaning of the result obtained.

**9.3.4** A modification of the equations described in exercise 9.3.3 is given by the following $SIS$ model,

$$\frac{dS}{dt} = -\frac{\beta}{N} S I + \gamma I, \qquad \frac{dI}{dt} = \frac{\beta}{N} S I - \gamma I.$$

Here, $\beta$ and $\gamma$ are positive constants of proportionality. We have again the relation $N = S + I$.

a) Explain the relevance of a nonzero $\gamma$.

b) Follow the approach in exercise 9.3.3 to derive a closed equation for $I$.

c) Solve the differential equation for $I$.

d) Calculate $S$ and $I$ for large values of $t$ for the cases that $\beta > \gamma$ and $\beta \leq \gamma$, respectively.

**9.3.5** A modification of the equations described in exercise 9.3.3 is given by the following $SIR$ model,

$$\frac{dS}{dt} = -\frac{\beta}{N} S I, \qquad \frac{dI}{dt} = \frac{\beta}{N} S I - \gamma I, \qquad \frac{dR}{dt} = \gamma I.$$

Here, $\beta$ and $\gamma$ are positive constants of proportionality. This model implies the relation $N = S(t) + I(t) + R(t)$.

a) Explain the difference between the assumptions reflected in this model and the *SIS* model.
b) Due to the relation $N = S(t) + I(t) + R(t)$ we can focus on the dynamics of $S(t)$ and $I(t)$. Use the model considered to derive an equation for the derivative $dI / dS$.
c) Use this equation to calculate $I = I(S)$. Rewrite this equation by using the abbreviations $y = \beta(I - I_0) / (\gamma N)$, $x = 1 - S / S_0$, and $R = \beta S_0 / (\gamma N)$. Here, $I_0$ and $S_0$ refer to the initial values of $I$ and $S$, respectively.
d) An epidemic occurs if $y = y(x)$ increases from its initial value zero to a local maximum. Under which condition can this happen? Determine the critical point of $x$ and the maximum of $y$. Hint: consider the fact that $0 \le x \le 1$ because $S$ is a decreasing function by $dS / dt = -\beta S I / N$.
e) No epidemic occurs if $y$ is a decreasing function of $x$. Show under which condition this is the case.

**9.3.6** A modification of the equations described in exercise 9.3.3 is given by the following *SIRS* model, which is given by the equations

$$\frac{dS}{dt} = -\frac{\beta}{N} S I + \nu R, \qquad \frac{dI}{dt} = \frac{\beta}{N} S I - \gamma I, \qquad \frac{dR}{dt} = \gamma I - \nu R.$$

Here, $\beta$, $\gamma$, and $\nu$ are positive constants of proportionality. This model implies the relation $N = S(t) + I(t) + R(t)$.
a) We can focus on the dynamics of $S(t)$ and $I(t)$ because $R$ is determined via the relation $N = S(t) + I(t) + R(t)$. Use the *SIRS* model considered to derive a closed equation system for $S^*(t) = N - S(t)$ and $I(t)$.
b) Determine the equilibrium points implied by the equations for $S^*$ and $I$.
c) Under which condition for the model parameters $\beta$, $\gamma$, and $\nu$ involved do we find positive equilibrium values for $S(t)$ and $I(t)$?
d) Consider the case that the parameter condition derived in c) is satisfied. Which of the equilibrium points will be realized?

**9.4.1** Consider the total energy $E = m\,g\,r\,(1 - \cos\alpha) + m\,r^2\,(d\alpha / dt)^2 / 2$. Use the fact that $E$ is constant for the undamped pendulum to derive the differential equation for the undamped pendulum. Hint: differentiate $E$.

**9.4.2** The undamped nonlinear pendulum equation $d^2\alpha / dt_*^2 + \sin\alpha = 0$ combined with the initial conditions $\alpha(0) = \alpha_0$ and $d\alpha / dt_*(0) = 0$ can be used to find an exact expression for the pendulum period $T_P$, which is the time required for the pendulum bob to swing through one complete cycle and return to its original position. This expression for $T_P$ reads (Boyce & DiPrima 2009)

$$T_P = 4\sqrt{\frac{r}{g}} \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - \sin^2(\alpha_0 / 2)\,\sin^2\theta}}.$$

The integral represents an elliptic integral of the first kind.

a) Consider the case that the initial angle $\alpha_0$ is very small. Calculate the integral by using the approximation $(1-x)^{-1/2} = 1 + x/2$. Hint: use the integral $\int \sin^2 x\, dx = x/2 - (1/4)\sin 2x$.

b) For which initial angles $\alpha_0$ is the influence of $\alpha_0$ on $T_P$ smaller than 1%?

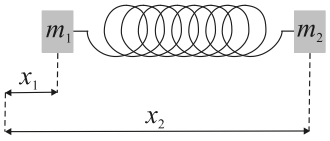
**9.4.3** Consider the undamped nonlinear pendulum equation (9.142), this means $y_2 = \pm\, 2^{1/2}\,(E_* + \cos y_1 - 1)^{1/2}$.

a) Write this formula in an explicit dependence on any initial conditions $y_{10}$ and $y_{20}$.

b) Explain under which conditions the positive and negative signs in the formula for $y_2$ have to be used, respectively.

c) For which $y_{20}$ do we always find open curves in the $y_1$-$y_2$ phase plane?

**9.4.4** We consider a spring-mass system that involves two coupled masses $m_1$ and $m_2$: see the corresponding illustration. The masses can move in one direction. Their positions are $x_1$ and $x_2$. According to Newton's Second Law, this spring-mass system can be described by the equation system (Haberman 1977)




$$m_1 \frac{d^2 x_1}{dt^2} = k\,(x_2 - x_1 - L), \qquad m_2 \frac{d^2 x_2}{dt^2} = -k\,(x_2 - x_1 - L).$$

Here, $k$ is the spring constant, and $L$ is the unstreched length of the spring.

a) The center of mass is defined by $z = (m_1 x_1 + m_2 x_2)/(m_1 + m_2)$. Find an equation for $z$ and solve it.

b) Consider the spring stretching $y = x_2 - x_1 - L$. Derive an equation for $y$.

c) Compare the $y$ equation with the undamped spring-mass system equation $d^2y/dt^2 + k\,y/m = 0$: see Eq. (7.42). Explain your observations.

d) Solve the $y$ equation. Hint: use the results derived in Sect. 7.3.3.

**9.4.5** Consider again the spring-mass system given in exercise 9.4.4.

a) Solve the equations for $x_1(t)$ and $x_2(t)$ for initial conditions chosen such that the initial values of the derivatives of $z$ and $y$ are zero ($z'_0 = y'_0 = 0$) and initial values of $z$ and $y$ that are given by $z_0 = 0$ and $y_0 = 1$. Hint: use the $y$ and $z$ solutions derived in exercise 9.4.4.

b) Calculate the positions $x_1(t)$ and $x_2(t)$ for the case that $m_1 \to \infty$. Explain why the spring-mass system motion obtained in this way makes sense.

c) Calculate the positions $x_1(t)$ and $x_2(t)$ for the case that $m_1 \to 0$. Explain why the spring-mass system motion obtained in this way makes sense.

**9.5.1** Consider the Lorenz equations (9.144).

a) Calculate the equilibrium points by accounting for variable values of the model parameters $Pr$ and $b$.

b) Show that the characteristic number $R_2$, which separates nonchaotic and chaotic solutions of the Lorenz equations, is given by

$$R_2 = Pr\frac{Pr+b+3}{Pr-b-1}.$$

Hint: follow the explanations in Sect. 9.5.2.

**9.5.2** Consider the expression for $R_2$ given in exercise 9.5.1.

a) Show the effect of increasing values of $b$ on $R_2$.

b) For which $Pr$ values is $R_2$ positive?

c) Consider $Pr$ values so that $R_2$ is positive. For growing $Pr$ values, $R_2$ decreases, it attains a minimum, and it increases. Find the minimum of $R_2$.

**9.5.3** O. Rössler (1976) analyzed the following equation system ($b$ represents a non-negative model parameter)

$$\frac{dy_1}{dt} = -y_2 - y_3, \qquad \frac{dy_2}{dt} = y_1 + 0.2\,y_2, \qquad \frac{dy_3}{dt} = 0.2 + y_3\,(y_1 - b).$$

a) Show that there exist two potential equilibrium points, which are given by $P = (0.2\,Y_3, -Y_3, Y_3)$. Here, $Y_3 = (5/2)\,[b \pm (b^2 - 4/25)^{1/2}]$.

b) What are the conditions to have no equilibrium point, one equilibrium point, and two equilibrium points?

c) Which behavior of solutions do you expect for the case that there is no equilibrium point?

**9.5.4** Consider the Rössler equations given in exercise 9.5.3.

a) Determine the linear equation system near the equilibrium points.

b) Show that the characteristic equation, which characterizes the behavior of the linear equation system obtained in a), is given by

$$0 = r^3 + 0.5\left(b - 0.4 \mp \sqrt{b^2 - 4/25}\right)r^2 + \left(2.4b + 1 \pm 2.6\sqrt{b^2 - 4/25}\right)r \mp \sqrt{b^2 - 4/25}.$$

The upper (lower) sign in this equation refers to the positive (negative) sign in $Y_3 = (5/2)\,[b \pm (b^2 - 4/25)^{1/2}]$.

c) Show that the condition for the critical value of $b$ that implies two pure imaginary eigenvalues is given by the equation

$$0 = \sqrt{b - 0.4}\left(\sqrt{b(b^2 - 0.12) - 0.016} \mp \sqrt{b(b^2 - 0.12) + 0.016}\right).$$

Hint: follow the explanations in Sect. 9.5.2.

d) What is the conclusion of the latter equation regarding the critical value of $b$ that implies two pure imaginary eigenvalues?

e) It was found for the Lorenz equations that the critical value of $b$ that implies two pure imaginary eigenvalues separates complex eigenvalues with positive and negative real parts (nonchaotic and chaotic solutions of the Lorenz equations). Does the critical value of $b$ determined in d) have the same property?

**9.5.5** Consider the Rössler equations in exercise 9.5.3 combined with $b = 0.4$.

a) Use the results given in exercises 9.5.3 and 9.5.4 to find the equilibrium point and the roots $r$ of the characteristic equation. Explain the stability behavior of solutions near the equilibrium point.

b) Show the validity of the findings obtained in a) in terms of $y_1$-$y_2$, $y_1$-$y_3$, and $y_2$-$y_3$ phase plane plots. Solve the Rössler equations up to $t = 40$ to obtain these figures. Use the initial values $(y_{10}, y_{20}, y_{30}) = (0.25, -1.05, 0.95)$ and $\Delta t = 10^{-4}$ for the numerical solution corresponding to the numerical scheme (9.172) used for the solution of the Lorenz equations.

**9.5.6** Consider the Rössler equations in exercise 9.5.3 combined with $b = 0.5$.

a) Use the results given in exercises 9.5.3 and 9.5.4 to find the first equilibrium point and the related roots $r$ of the characteristic equation. What do the results obtained mean regarding the stability behavior of solutions near the first equilibrium point?

b) Show the validity of the findings obtained in a) by $y_1$-$y_2$, $y_1$-$y_3$, and $y_2$-$y_3$ phase plane plots. Solve the Rössler equations up to $t = 40$ to obtain these figures. Use the initial values $(y_{10}, y_{20}, y_{30}) = (0.35, -2.1, 1.9)$ and $\Delta t = 10^{-4}$ for the numerical solution corresponding to the numerical scheme (9.172) used for the solution of the Lorenz equations.

c) Find the second equilibrium point and the related roots $r$ of the characteristic equation. Explain the meaning of the results obtained regarding the stability behavior of solutions near the second equilibrium point.

d) Show the validity of the findings obtained in c) by means of $y_1$-$y_2$, $y_1$-$y_3$, and $y_2$-$y_3$ phase plane plots. Solve the Rössler equations up to $t = 100$ to obtain these figures. Use the initial values $(y_{10}, y_{20}, y_{30}) = (0, -0.4, 0.4)$ and $\Delta t = 10^{-4}$ for the numerical solution.

# 10 Stochastic Multivariate Evolution

The discussions of stochastic methods in previous chapters were related to the consideration of a single random variable. This approach is appropriate to explain the basic structure of evolution equations for stochastic processes and their PDFs. On the other hand, most applications cannot be handled on the basis of methods that describe the evolution of single variables. Real processes usually take place in the three-dimensional physical space, and they often involve several variables. Examples are given by flow phenomena (the three-dimensional atmospheric wind field that interacts with the temperature), chemical reactor processes (involving a variety of chemical species in three-dimensional reactors), and the competition of several population densities in areas with varying food resources. To prepare the application of stochastic methods to such cases we will extend now the methods developed in Chap. 8 to the case of several random variables. In fact, the methods to be described in this chapter are applicable to a wide range of realistic problems. More detailed descriptions of corresponding applications can be found elsewhere (Pope 2000, Roekaerts 2002, Heinz 2003, Fox 2003, Givi 2006). From a mathematical point of view, the discussion here reveals a relationship between partial differential equations and stochastic ordinary differential equations, which is very helpful for the solution of complicated partial differential equations.

Section 10.1 explains the motivation for considering joint processes of several random variables. Joint PDFs that do not evolve will be considered in Sects. 10.2 and 10.3: Sect. 10.2 explains the definition of joint PDFs and Sect. 10.3 presents the normal model for joint PDFs. Joint PDFs that evolve in time will be considered in Sects. 10.4 and 10.5. The concepts for the description of the evolution of a single-variable PDF (and the corresponding stochastic process) will be extended to the several-variable case in Sect. 10.4. Section 10.5 explains the application of such equations to the modeling of molecular and fluid motion. Section 10.6 summarizes the basic observations made in this chapter.

S. Heinz, *Mathematical Modeling*, DOI 10.1007/978-3-642-20311-4_10,

## 10.1 Motivation

**Fluid Dynamics.** As an example, let us consider the motion of fluids (e.g., atmospheric motions) in order to illustrate the need for methods for the calculation of the evolution of several random variables. The prediction of fluid flow requires the calculation of the mean velocity $U_i(\boldsymbol{x}, t)$ of molecules, which represents the $i^{\text{th}}$ component ($i = 1, 3$) of the fluid velocity at the position $\boldsymbol{x} = (x_1, x_2, x_3)$ at time $t$. It will be shown in Sect. 10.5 that the fluid velocity $U_i(\boldsymbol{x}, t)$ and fluid mass density $\rho(\boldsymbol{x}, t)$ have to satisfy a coupled system of partial differential equations, which represent the conservation of mass and momentum,

$$\frac{D\rho}{Dt} + \rho \frac{\partial U_m}{\partial x_m} = 0, \tag{10.1a}$$

$$\frac{DU_i}{Dt} + \frac{1}{\rho} \frac{\partial \rho \sigma_{im}}{\partial x_m} = 0. \tag{10.1b}$$

Here, $\sigma_{im}(\boldsymbol{x}, t)$ refers to the variance of molecular velocities, this means $\sigma_{im} = \overline{v_i v_m}$ (see Sect. 10.5). We use the sum convention for repeated subscripts, this means we have for example

$$\frac{\partial U_m}{\partial x_m} = \frac{\partial U_1}{\partial x_1} + \frac{\partial U_2}{\partial x_2} + \frac{\partial U_3}{\partial x_3}. \tag{10.2}$$

The total derivative (or substantial or material derivative) of any property $Q(\boldsymbol{x}, t)$ (we may set, for example, $Q = \rho$ or $Q = U_i$) is defined by

$$\frac{DQ}{Dt} = \frac{\partial Q}{\partial t} + U_m \frac{\partial Q}{\partial x_m}. \tag{10.3}$$

The meaning of $DQ / Dt$ can be seen by considering the property $Q$ at $\boldsymbol{x} = \boldsymbol{x}(t)$. Here, $\boldsymbol{x}(t)$ is a point that follows the fluid velocity $U_i$, i.e., $\boldsymbol{x}(t)$ is determined by

$$\frac{dx_i(t)}{dt} = U_i\big(\boldsymbol{x}(t), t\big). \tag{10.4}$$

The total derivative $DQ / Dt$ at $\boldsymbol{x} = \boldsymbol{x}(t)$ reads

$$\begin{aligned} \frac{DQ\big(\boldsymbol{x}(t), t\big)}{Dt} &= \frac{\partial Q\big(\boldsymbol{x}(t), t\big)}{\partial t} + U_m\big(\boldsymbol{x}(t), t\big) \frac{\partial Q\big(\boldsymbol{x}(t), t\big)}{\partial x_m} \\ &= \frac{\partial Q\big(\boldsymbol{x}(t), t\big)}{\partial t} + \frac{\partial Q\big(\boldsymbol{x}(t), t\big)}{\partial x_m} \frac{dx_m(t)}{dt} = \frac{dQ\big(\boldsymbol{x}(t), t\big)}{dt}. \end{aligned} \tag{10.5}$$

The last line makes use of Eq. (10.4). Hence, $DQ / Dt$ represents the total change of the property $Q$ in time at a point $\boldsymbol{x}(t)$ moving with the fluid velocity $U_i$.

**Closure Problem.** Equations (10.1) are unclosed because the variance $\sigma_{im}$ of molecular velocities is unknown. This is not a minor problem, but $\sigma_{im}$ determines the velocity change $DU_i/Dt$ according to Eq. (10.1b). The variance $\sigma_{im}$ has to satisfy a conservation equation, too (see Sect. 10.5),

$$\frac{D\sigma_{ij}}{Dt} + \frac{1}{\rho}\frac{\partial \rho \overline{v_i v_j v_m}}{\partial x_m} + \frac{\partial U_i}{\partial x_m}\sigma_{mj} + \frac{\partial U_j}{\partial x_m}\sigma_{mi} = -\frac{2}{T}\left(\sigma_{ij} - \frac{\sigma_{kk}}{3}\delta_{ij}\right). \tag{10.6}$$

Here, $\overline{v_i v_j v_m}$ is the triple correlation of molecular velocities, $T$ is a characteristic relaxation time scale, and $\delta_{ij}$ refers to the Kronecker delta (which is zero for $i \neq j$ and one for $i = j$). This equation is again unclosed because the triple correlation is unknown. It would be possible to continue in this way by considering an equation for the triple correlation. However, this equation does again contain an unknown correlation of higher order, and this applies to all such equations. The solution of this closure problem requires a model that explains the evolution of all moments of molecular velocities, which define the joint PDF of the three molecular velocity components. Therefore, we need a model for the evolution of this joint PDF. Such a model will be presented in Sect. 10.5.

**Questions Considered.** Hence, we have to extend the methods for the analysis and modeling of single random variables to the description of properties of several random variables. In particular, we need answers to the following questions:

- How can we extend concepts for the data analysis of single random variables to concepts for the data analysis of joint random variables?
- How can we extend usual PDF models for single random variables (for example, the normal PDF model) to the case of several variables?
- How can we extend PDF equations for the description of the evolution of single-variable PDFs to equations for the evolution of joint PDF of several variables?

The first two questions will be considered in Sects. 10.2 and 10.3 for the case of two random variables by focusing on the data analysis. The last question will be addressed in Sects. 10.4 and 10.5 with focus on the modeling of several-variable processes.

## 10.2 Data Analysis Concepts for Joint Random Variables

How can we extend concepts for the data analysis of single random variables to concepts for the data analysis of joint random variables? First of all, this requires the definition of a joint PDF, this means the explanation of how a joint PDF can be obtained from measurements. It will be also helpful to extend the definitions of a single-variable PDF and its moments introduced in Chap. 4 by the consideration of correlations, which corresponds to the introduction of conditional means. These

questions will be addressed in this section by considering PDFs of two random variables $X$ and $Y$, which may have values between negative and positive infinity. The concepts to be developed can be straightforwardly extended to the case of many variables. Such multidimensional joint PDFs will be considered in Sect. 10.4.1 in the context of the discussion of evolution equations for joint PDFs.

### *10.2.1 Joint Probability Density Functions*

**Joint PDF.** In extension of the definition $f(x) = \langle\delta(x - X)\rangle$ of the PDF of a single random variable $X$, we define the joint PDF of two variables $X$ and $Y$ by

$$f(x,y) = \Big\langle \delta(x-X)\,\delta(y-Y) \Big\rangle. \tag{10.7}$$

The joint PDF $f(x, y)$ has the properties

$$\int f(x,y)\,dy = \int \Big\langle \delta(x-X)\,\delta(y-Y) \Big\rangle\,dy = \Big\langle \delta(x-X) \Big\rangle = f(x). \tag{10.8a}$$

$$\int f(x,y)\,dx = \int \Big\langle \delta(x-X)\,\delta(y-Y) \Big\rangle\,dx = \Big\langle \delta(y-Y) \Big\rangle = f(y). \tag{10.8b}$$

The first rewriting of the left-hand sides makes use of the definition (10.7) of the joint PDF $f(x, y)$. The second rewriting applies the normalization property of delta functions. The PDFs $f(x)$ and $f(y)$ of single variables are called marginal PDFs. As shown in exercise 10.2.1, other typical properties of the joint PDF $f(x, y)$ are

$$f(x,y) \geq 0, \tag{10.9a}$$

$$f(-\infty,y) = f(\infty,y) = f(x,-\infty) = f(x,\infty) = 0, \tag{10.9b}$$

$$\iint f(x,y)\,dx\,dy = 1, \tag{10.9c}$$

$$\iint g(x,y)\; f(x,y)\,dx\,dy = \Big\langle g(X,Y) \Big\rangle, \tag{10.9d}$$

where $g(x, y)$ is any function of $x$ and $y$. The knowledge of the joint PDF $f(x, y)$ enables the calculation of the probability for joint events $a \leq X \leq b$ and $c \leq Y \leq d$,

$$P(a \leq X \leq b,\, c \leq Y \leq d) = \int_c^d \left[ \int_a^b f(x,y)\,dx \right] dy. \tag{10.10}$$

The validity of this relation can be seen by using the definition (10.7) of $f(x, y)$,

$$\begin{aligned} \int_c^d \left[ \int_a^b f(x,y)\,dx \right] dy &= \left\langle \int_c^d \left[ \int_a^b \frac{d\theta(x-X)}{dx}\,\frac{d\theta(y-Y)}{dy}\,dx \right] dy \right\rangle \\ &= \Big\langle \big(\theta(b-X) - \theta(a-X)\big)\big(\theta(d-Y) - \theta(c-Y)\big) \Big\rangle. \end{aligned} \tag{10.11}$$

Here, the delta functions were replaced by derivatives of theta functions according to $\delta(x-X) = d\,\theta(x-X)/dx$. Relation (10.10) can be specified for the case

$$P(x \le X \le x+dx, y \le Y \le y+dy) = \int_{y}^{y+dy} \left[ \int_{x}^{x+dx} f(\hat{x},\hat{y})\, d\hat{x} \right] d\hat{y}, \tag{10.12}$$

where $dx$ and $dy$ are infinitesimal intervals. In the first order of approximation we can replace $f(\hat{x},\hat{y})$ in the integral by $f(x, y)$. Then, Eq. (10.12) provides

$$P(x \le X \le x+dx, y \le Y \le y+dy) = f(x,y)\, dx\, dy. \tag{10.13}$$

Hence, $f(x, y)$ determines the probability to find $X$ and $Y$ in infinitesimal intervals at $x$ and $y$.

**Independence.** The joint PDF $f(x, y)$ becomes simpler for the specific case of independent random variables, this means for the case that there is no effect of one variable on the other variable. For this case, the joint PDF can be written

$$f(x,y) = \big\langle \delta(x-X) \big\rangle \big\langle \delta(y-Y) \big\rangle = f(x)\, f(y). \tag{10.14}$$

The consideration of independent variables simplifies analyses significantly. The concepts of independent and uncorrelated random variables (variables with a zero correlation coefficient $r_{XY}$, see Sect. 2.3.1) are similar but different. Independent variables are always uncorrelated: the correlation coefficient $r_{XY} = 0$. However, the converse is not true in general: uncorrelated variables do not have to be independent. An example for the latter case is the following: Let $X$ be uniformly distributed on [−1, 1] and $Y = X^2$. The calculation of the correlation coefficient then shows that both variables are uncorrelated. However, $X$ determines $Y$, and $Y$ restricts $X$ to at most two values. Hence, $X$ and $Y$ are not independent variables.

### *10.2.2 Conditional Probability Density Functions*

**Conditional PDF.** The joint PDF $f(x, y)$ determines the probability to find $X$ and $Y$ in infinitesimal intervals at $x$ and $y$. However, there is relatively often a slightly different problem given by the question of what is the probability to find values of one variable (e.g., $y$) for a fixed value of the other variable (e.g., $x$): see the discussion in Sect. 10.2.3. Information regarding this question is given by the joint PDF $f(x, y)$, but $f(x, y)$ does not represent a PDF for $y$ (the integral over $y$ does not result in one: see Eq. (10.8a)). Therefore, the joint PDF is rescaled so that the rescaled PDF integrates to one. This rescaled PDF is given by

$$f(y \mid x) = \frac{f(x,y)}{f(x)}. \tag{10.15}$$

The PDF $f(y \mid x)$ is called the PDF of $y$ conditioned on $x$ (or simply conditional PDF): it describes the probability to find $y$ values under the condition that $X = x$. The integral of the conditional PDF $f(y \mid x)$ over $y$ is equal to one,

$$\int f(y \mid x)\, dy = 1, \tag{10.16}$$

which follows from $f(y \mid x) = f(x, y) / f(x)$ and the property (10.8a) of $f(x, y)$. The conditional PDF of independent variables, for which we have $f(x, y) = f(x) f(y)$, is equal to the corresponding unconditional PDF, $f(y \mid x) = f(y)$.

**Conditional Mean.** The conditional PDF can be used to define a conditional mean. With regard to any function $g(x, y)$, this relation reads

$$\int g(x,y)\, f(y \mid x)\, dy = \big\langle g(X,Y) \mid x \big\rangle. \tag{10.17}$$

By using the definition $f(y \mid x) = f(x, y) / f(x)$, this relation also can be written

$$\int g(x,y)\, f(x,y)\, dy = \big\langle g(X,Y) \mid x \big\rangle\, f(x). \tag{10.18}$$

The consistency of this relation can be seen by integrating it over $x$,

$$\int \big\langle g(X,Y) \mid x \big\rangle\, f(x)\, dx = \int\!\!\int g(x,y)\, f(x,y)\, dy\, dx = \big\langle g(X,Y) \big\rangle. \tag{10.19}$$

The last expression follows from the property (10.9d) of joint PDFs. Hence, the integral over the conditional mean multiplied with the probability to find $x$ equals the unconditional mean.

**Conditional Mean Calculation.** Equation (10.17) can be used to calculate a conditional mean, but this requires the joint PDF $f(x, y)$ and the integration over $y$. This can be avoided by performing the integration over $y$ in Eq. (10.17),

$$\begin{aligned} \big\langle g(X,Y) \mid x \big\rangle &= \frac{1}{f(x)} \int g(x,y) \big\langle \delta(x - X)\, \delta(y - Y) \big\rangle\, dy \\ &= \frac{1}{f(x)} \Big\langle g(X,Y) \int \delta(x - X)\, \delta(y - Y)\, dy \Big\rangle = \frac{1}{f(x)} \big\langle g(X,Y)\, \delta(x - X) \big\rangle. \end{aligned} \tag{10.20}$$

The first rewriting of the conditional mean is obtained by replacing the joint PDF $f(x, y)$ in the conditional PDF $f(y \mid x) = f(x, y) / f(x)$ by its definition (10.7). In the second line, the mean value is used for all the integral, and the sifting property of delta functions is used, such that $g(X, Y)$ can be written in front of the integral. The last expression results from the normalization condition for delta functions. The relation between a conditional mean and a conditional PDF can be seen by setting $g(X, Y) = \delta(y - Y)$. For this case, Eq. (10.20) will become

$$\big\langle \delta(y - Y) \mid x \big\rangle = \frac{1}{f(x)} \big\langle \delta(y - Y)\, \delta(x - X) \big\rangle = \frac{f(x,y)}{f(x)} = f(y \mid x). \tag{10.21}$$

These rewritings follow from the definitions of a joint PDF $f(x, y)$ and conditional PDF $f(y \mid x)$. Thus, the conditional PDF represents a conditional mean.

### *10.2.3 Application to Optimal Modeling*

**Optimal Models.** Typical problems involving two random variables were considered in Chap. 2. We considered a set of $(X_i, Y_i)$ data, where $i = 1, N$. The problem was to find a model $y_M(x)$ that agrees as good as possible with the given data. The particular problem was to find a model $y_M(x)$ that minimizes the least-squares error

$$E^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - y_M(X_i)\right)^2. \tag{10.22}$$

The objective here is not to make any modifications of the approach presented in Chap. 2, but to present the findings obtained in Chap. 2 in terms of properties of random variables.

**Error Definition.** In terms of the notation applied here, the least-squares error can be written as a mean value,

$$E^2 = \left\langle \left[Y - y_M(X)\right]^2 \right\rangle. \tag{10.23}$$

According to Eq. (10.19), the least-squares error $E^2$ also can be written in terms of a conditional mean,

$$E^2 = \int \left\langle \left[Y - y_M(X)\right]^2 \mid x \right\rangle f(x)\, dx = \int \left\langle \left[Y - y_M(x)\right]^2 \mid x \right\rangle f(x)\, dx. \tag{10.24}$$

The last expression accounts for the condition $X = x$. The advantage of using the conditional mean is that the error $E^2$ is now related to the function $y_M(x)$, which has to be calculated.

**Minimal Error.** Which model function $y_M(x)$ could minimize the least-squares error? The last expression in Eq. (10.24) represents the mean value of the non-negative numbers $<[Y - y_M(x)]^2 \mid x>$. Therefore, the minimum value of $E^2$ is given if $<[Y - y_M(x)]^2 \mid x>$ becomes minimal. This conditional mean can be written

$$\begin{aligned} \left\langle \left[Y - y_M(x)\right]^2 \mid x \right\rangle &= \left\langle \left[Y - \langle Y \mid x\rangle + \langle Y \mid x\rangle - y_M(x)\right]^2 \mid x \right\rangle \\ &= \left\langle \left[Y - \langle Y \mid x\rangle\right]^2 \mid x \right\rangle + \left\langle \left[\langle Y \mid x\rangle - y_M(x)\right]^2 \mid x \right\rangle + h(x). \end{aligned} \tag{10.25}$$

The first rewriting involves $<Y \mid x> - <Y \mid x>$. The second rewriting results from distributing the quadratic term, where the function $h(x)$ is given by

$$h(x) = 2\left\langle \left[Y - \langle Y \mid x\rangle\right]\left[\langle Y \mid x\rangle - y_M(x)\right] \mid x \right\rangle. \tag{10.26}$$

A closer look at $h(x)$ shows that $h(x) = 0$,

$$\begin{aligned} h(x) &= 2\Big[\big\langle Y \mid x\big\rangle - y_M(x)\Big]\big\langle\Big[Y - \big\langle Y \mid x\big\rangle\Big] \mid x\big\rangle \\ &= 2\Big[\big\langle Y \mid x\big\rangle - y_M(x)\Big]\Big[\big\langle Y \mid x\big\rangle - \big\langle Y \mid x\big\rangle\Big] = 0. \end{aligned} \tag{10.27}$$

The expression $<Y \mid x> - y_M(x)$ is unaffected by the condition, which results in the first line. The next line follows from distributing the conditional mean. Therefore, Eq. (10.25) can be written

$$\Big\langle\Big[Y - y_M(x)\Big]^2 \mid x\Big\rangle = \Big\langle\Big[Y - \big\langle Y \mid x\big\rangle\Big]^2 \mid x\Big\rangle + \Big[\big\langle Y \mid x\big\rangle - y_M(x)\Big]^2. \tag{10.28}$$

The model function $y_M(x)$ does only affect the last term, which is non-negative. Thus, the conditional mean $<[Y - y_M(x)]^2 \mid x>$ becomes minimal if the last term disappears, this means if

$$y_M(x) = \big\langle Y \mid x\big\rangle. \tag{10.29}$$

This expression is relevant: (*i*) it explains how an optimal model function $y_M(x)$ can be calculated on the basis of measured data (without making use of any model assumptions), (*ii*) it provides a basis for the optimization of model function types considered (see the explanations in the next paragraph), and (*iii*) it enables the calculation of $y_M(x)$ on the basis of a model for the joint PDF $f(x, y)$ of $X$ and $Y$ (see Sect. 10.3).

**Optimal Linear Model.** The usual way to address optimization problems is the attempt to transform the data such that a linear model can be used,

$$y_M(x) = ax + b. \tag{10.30}$$

The model parameters $a$ and $b$ can be calculated in terms of Eq. (10.29). By using $y_M(x) = <Y \mid x>$ and the definition (10.20) of conditional means, Eq. (10.30) multiplied by $f(x)$ can be written

$$\big\langle Y\,\delta(x - X)\big\rangle = \big(ax + b\big) f(x). \tag{10.31}$$

We take the integral over $x$ to obtain a condition for $b$,

$$\big\langle Y\big\rangle = a\big\langle X\big\rangle + b. \tag{10.32}$$

By replacing the parameter $b$ in Eq. (10.31) by this condition we obtain

$$\big\langle Y\,\delta(x - X)\big\rangle = \big\langle Y\big\rangle f(x) + a\Big(x - \big\langle X\big\rangle\Big) f(x). \tag{10.33}$$

The term $<Y> f(x)$ can be combined with the left-hand side,

$$\Big\langle \widetilde{Y}\,\delta(x - X)\Big\rangle = a\Big(x - \big\langle X\big\rangle\Big) f(x). \tag{10.34}$$

The multiplication of this expression by $x - \langle X \rangle$ and integration over $x$ then provides a condition for the model parameter $a$,

$$\begin{aligned} a\left\langle \widetilde{X}^2 \right\rangle &= \left\langle \int \left(x - \left\langle X \right\rangle\right) \widetilde{Y}\, \delta(x - X)\, dx \right\rangle = \left\langle \int \left(X - \left\langle X \right\rangle\right) \widetilde{Y}\, \delta(x - X)\, dx \right\rangle \\ &= \left\langle \left(X - \left\langle X \right\rangle\right) \widetilde{Y} \int \delta(x - X)\, dx \right\rangle = \left\langle \widetilde{X}\, \widetilde{Y} \right\rangle. \end{aligned} \tag{10.35}$$

Here, we used the sifting property and normalization condition of delta functions. By combing $y_M(x) = \langle Y \,|\, x \rangle = a\,x + b$ with Eqs. (10.32) and (10.35) we get

$$y_M(x) = \left\langle Y \mid x \right\rangle = \left\langle Y \right\rangle + \frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle}\left(x - \left\langle X \right\rangle\right) = \left\langle Y \right\rangle + r_{XY} \left\langle \widetilde{Y}^2 \right\rangle^{1/2} \frac{x - \left\langle X \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}}. \tag{10.36}$$

This result recovers Eq. (2.47). The last expression applies the correlation coefficient, which was already defined in Chap. 2,

$$r_{XY} = \frac{\left\langle \widetilde{X}\, \widetilde{Y} \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle^{1/2} \left\langle \widetilde{Y}^2 \right\rangle^{1/2}}. \tag{10.37}$$

The use of the latter expression for $y_M(x)$ in Eq. (10.23) results in the following minimal least-squares error

$$\begin{aligned} E^2 &= \left\langle \left[Y - y_M(X)\right]^2 \right\rangle = \left\langle \left[\widetilde{Y} - \widetilde{X}\left\langle \widetilde{X}\widetilde{Y} \right\rangle / \left\langle \widetilde{X}^2 \right\rangle\right]^2 \right\rangle \\ &= \left\langle \widetilde{Y}^2 \right\rangle - 2\frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle}{\left\langle \widetilde{X}^2 \right\rangle}\left\langle \widetilde{X}\widetilde{Y} \right\rangle + \frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle^2}{\left\langle \widetilde{X}^2 \right\rangle^2}\left\langle \widetilde{X}^2 \right\rangle \\ &= \left\langle \widetilde{Y}^2 \right\rangle - \frac{\left\langle \widetilde{X}\widetilde{Y} \right\rangle^2}{\left\langle \widetilde{X}^2 \right\rangle} = \left\langle \widetilde{Y}^2 \right\rangle (1 - r_{XY}{}^2). \end{aligned} \tag{10.38}$$

This expression recovers Eq. (2.53) for $E^2$, see the related discussion in Chap. 2.

## 10.3 The Joint Normal Pobability Density Function Model

Let us address now the question of how joint PDFs can be modeled. We will consider here the extension of the normal PDF model for single variables, which represents the most relevant PDF model for unbounded variables, to the case of two correlated random variables. The extension to the many-variable case will be described in the context of Fokker-Planck equations (see Sect. 10.4).

### *10.3.1 The Joint Normal Probability Density Function Model*

**Joint Normal PDF.** The joint normal PDF $f(x, y)$ of two random variables $X$ and $Y$ can be defined by

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi\sqrt{(1-r_{XY}{}^2)\left\langle\widetilde{X}^2\right\rangle\left\langle\widetilde{Y}^2\right\rangle}}\exp\left\{-\frac{\hat{x}^2+\hat{y}^2-2r_{XY}\hat{x}\,\hat{y}}{2(1-r_{XY}{}^2)}\right\}, \\ &= \frac{1}{2\pi\sqrt{(1-r_{XY}{}^2)\left\langle\widetilde{X}^2\right\rangle\left\langle\widetilde{Y}^2\right\rangle}}\exp\left\{-\frac{(\hat{y}-r_{XY}\hat{x})^2+(1-r_{XY}{}^2)\,\hat{x}^2}{2(1-r_{XY}{}^2)}\right\}. \end{aligned} \tag{10.39}$$

The second line represents a convenient rewriting of the first line, which will be used below. To represent these expressions efficiently we applied here the non-dimensional variables

$$\hat{x} = \frac{x-\langle X\rangle}{\left\langle\widetilde{X}^2\right\rangle^{1/2}}, \qquad \hat{y} = \frac{y-\langle Y\rangle}{\left\langle\widetilde{Y}^2\right\rangle^{1/2}} \tag{10.40}$$

as abbreviations. The correlation coefficient $r_{XY}$ is given by Eq. (10.37). By defining normalized random variables

$$\hat{X} = \frac{\widetilde{X}}{\left\langle\widetilde{X}^2\right\rangle^{1/2}}, \qquad \hat{Y} = \frac{\widetilde{Y}}{\left\langle\widetilde{Y}^2\right\rangle^{1/2}} \tag{10.41}$$

in analogy to Eqs. (10.40), we find the correlation coefficient $r_{XY}$ to be given by

$$r_{XY} = \left\langle\hat{X}\,\hat{Y}\right\rangle. \tag{10.42}$$

Due to $|r_{XY}| \le 1$ we have $1 - r_{XY}{}^2 \ge 0$, i.e., the variance in Eqs. (10.39) is non-negative. The model (10.39) does satisfy the consistency conditions (10.8), see exercise 10.3.1.

**Moments.** An efficient way to present the moments of the joint PDF $f(x, y)$ is to do this in terms of the normalized random variables (10.41). The moments can be calculated by multiplying the PDF $f(x, y)$ with the corresponding variables and integration. Similar to the properties of a single-variable normal PDF it is found that the third-order and fifth-order (and all other odd-numbered) central moments are equal to zero,

$$\begin{aligned} &\left\langle\hat{X}^3\right\rangle = \left\langle\hat{X}^2\,\hat{Y}\right\rangle = \left\langle\hat{X}\,\hat{Y}^2\right\rangle = \left\langle\hat{Y}^3\right\rangle = 0, \\ &\left\langle\hat{X}^5\right\rangle = \left\langle\hat{X}^4\,\hat{Y}\right\rangle = \left\langle\hat{X}^3\,\hat{Y}^2\right\rangle = \left\langle\hat{X}^2\,\hat{Y}^3\right\rangle = \left\langle\hat{X}\,\hat{Y}^4\right\rangle = \left\langle\hat{Y}^5\right\rangle = 0. \end{aligned} \tag{10.43a}$$

The even-numbered normalized central moments are functions of the correlation coefficient $r_{XY}$. For example, the fourth-order and sixth-order central moments are given by

$$\begin{aligned}
&\left\langle \hat{X}^4 \right\rangle = \left\langle \hat{Y}^4 \right\rangle = 3, && \left\langle \hat{X}^6 \right\rangle = \left\langle \hat{Y}^6 \right\rangle = 15,\\
&\left\langle \hat{X}^3\,\hat{Y} \right\rangle = \left\langle \hat{X}\,\hat{Y}^3 \right\rangle = 3\,r_{XY}, && \left\langle \hat{X}^5\,\hat{Y} \right\rangle = \left\langle \hat{X}\,\hat{Y}^5 \right\rangle = 15\,r_{XY},\\
&\left\langle \hat{X}^2\,\hat{Y}^2 \right\rangle = 1 + 2\,{r_{XY}}^2, && \left\langle \hat{X}^4\,\hat{Y}^2 \right\rangle = \left\langle \hat{X}^2\,\hat{Y}^4 \right\rangle = 3 + 12\,{r_{XY}}^2,\\
& && \left\langle \hat{X}^3\,\hat{Y}^3 \right\rangle = 9\,r_{XY} + 6\,{r_{XY}}^3.
\end{aligned} \tag{10.43b}$$

For independent variables for which we have $r_{XY} = 0$, these relations recover the consequences for single normally distributed variables. Equations (10.43) can be used to decide whether any joint PDF is normal or not (see, e.g., the discussion of this question regarding the Brownian motion model in Sect. 6.4.2). For the case that all the conditions (10.43) implied by a joint normal PDF are satisfied, we can conclude that the joint PDF considered represents a normal PDF. Why is this conclusion valid? It is possible that another joint PDF implies moments that agree with some of the relations considered here (this PDF may also imply zero third-order and fifth-order moments), but it is impossible that another joint PDF implies moments that agree with all the 22 conditions (10.43).

**Conditional PDF.** Equation (10.39) can be used for writing the joint PDF as

$$f(x,y) = \frac{1}{\sqrt{2\pi(1-{r_{XY}}^2)\left\langle \widetilde{Y}^2 \right\rangle}} \exp\left\{ -\frac{(\hat{y} - r_{XY}\hat{x})^2}{2(1-{r_{XY}}^2)} \right\} f(x), \tag{10.44}$$

where $f(x)$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi\left\langle \widetilde{X}^2 \right\rangle}} \exp\left\{ -\frac{\hat{x}^2}{2} \right\}. \tag{10.45}$$

Comparison of Eq. (10.45) with the definition of the conditional PDF $f(y\,|\,x)$, i.e.,

$$f(x,y) = f(y\,|\,x) f(x), \tag{10.46}$$

shows that the conditional PDF $f(y\,|\,x)$ is given by

$$f(y\,|\,x) = \frac{1}{\sqrt{2\pi(1-{r_{XY}}^2)\left\langle \widetilde{Y}^2 \right\rangle}} \exp\left\{ -\frac{(\hat{y} - r_{XY}\hat{x})^2}{2(1-{r_{XY}}^2)} \right\}. \tag{10.47}$$

The conditional PDF $f(y\,|\,x)$ integrates to unity, $\int f(y\,|\,x)\,dy = 1$, see exercise 10.3.2. Considered as a function of $\hat{y}$, $f(y\,|\,x)$ represents a normal PDF with mean $r_{XY}\,\hat{x}$

and variance $1-r_{XY}^2$, which is divided by $<\widetilde{Y}^2>^{1/2}$. Hence, we have the relation

$$\int \hat{y}\, f(y \mid x)\, d\hat{y} = \frac{r_{XY}\hat{x}}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}}. \tag{10.48}$$

**Conditional Mean and PDF.** In terms of $f(y \mid x)$ we can obtain all conditional moments. First of all, we are interested in the conditional mean, which is given by

$$\left\langle Y \mid x \right\rangle = \int y\, f(y \mid x)\, dy. \tag{10.49}$$

The conditional mean can be calculated by writing Eq. (10.48) as

$$\int \frac{y - \left\langle Y \right\rangle}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}} f(y \mid x) \frac{dy}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}} = \frac{r_{XY}\hat{x}}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}}. \tag{10.50}$$

By using the definition (10.49) we find then for the conditional mean

$$\left\langle Y \mid x \right\rangle = \left\langle Y \right\rangle + r_{XY} \left\langle \widetilde{Y}^2 \right\rangle^{1/2} \hat{x} = \left\langle Y \right\rangle + r_{XY} \left\langle \widetilde{Y}^2 \right\rangle^{1/2} \frac{x - < X >}{\left\langle \widetilde{X}^2 \right\rangle^{1/2}}, \tag{10.51}$$

where $\hat{x}$ is used according to its definition (10.40). This expression for the conditional mean enables us to write the conditional PDF $f(y \mid x)$ given by Eq. (10.47) in a very convenient way. To prepare this representation we write

$$\hat{y} - r_{XY}\hat{x} = \frac{y - \left\langle Y \right\rangle}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}} - \frac{\left\langle Y \mid x \right\rangle - \left\langle Y \right\rangle}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}} = \frac{y - \left\langle Y \mid x \right\rangle}{\left\langle \widetilde{Y}^2 \right\rangle^{1/2}}, \tag{10.52}$$

where the definition of $\hat{y}$ and expression (10.51) for the conditional mean are applied. The use of this relation in Eq. (10.47) leads to the conclusion that

$$f(y \mid x) = \frac{1}{\sqrt{2\pi\,(1 - r_{XY}^2)\left\langle \widetilde{Y}^2 \right\rangle}} \exp\left\{ - \frac{\left(y - \left\langle Y \mid x \right\rangle\right)^2}{2(1 - r_{XY}^2)\left\langle \widetilde{Y}^2 \right\rangle} \right\}. \tag{10.53}$$

Therefore, the conditional PDF represents a normal PDF with mean $<Y \mid x>$ and variance $(1 - r_{XY}^2) <\widetilde{Y}^2>$. Thus, the deviations $Y - <Y \mid x>$ from the conditional mean are normally distributed with zero mean and variance $(1 - r_{XY}^2) <\widetilde{Y}^2>$, this means the deviations $Y - <Y \mid x>$ are independent of $x$.

**Statistical Formulation of Optimal Models.** In Sect. 10.2.3 we analyzed the consequences of considering a linear conditional mean, which leads to the global variance (10.38). Evidence for the suitability of considering such a mean and variance was not provided, which leads to the question of whether there is any conditional PDF that has such a mean and variance, and whether it is reasonable to

consider such a PDF. Answers to these questions are obtained by the findings obtained in the previous paragraph. We see that the assumption of a joint normal PDF, which is certainly reasonable, implies a linear conditional mean. The global variance (10.38) is also supported by this PDF, see exercise 10.3.3.

### *10.3.2 Data Analysis*

There is often the question of whether a joint normal PDF can be applied to model the joint PDF of given $X$ and $Y$ data. This question cannot be answered by considering only the marginal PDFs of $X$ and $Y$. For example, it is incorrect to conclude that two variables have a joint normal PDF even if the marginal PDFs of both $X$ and $Y$ are normal PDFs: the mixed moments may differ from Eq. (10.43b).

**Joint Normal PDF Features.** To address this question, it is helpful to know the characteristic features of a normal joint PDF. The best way to illustrate the joint PDF features is to consider isolines $f(x, y) = f$ in the $\hat{x} - \hat{y}$ plane, where $f$ is a constant. In this case, we can write Eq. (10.39) as

$$\begin{aligned}\hat{x}^2 + \hat{y}^2 - 2r_{XY}\hat{x}\,\hat{y} &= -2\,(1 - r_{XY}{}^2)\ln\left(2\pi\, f\sqrt{(1 - r_{XY}{}^2)\left\langle \widetilde{X}^2\right\rangle\left\langle \widetilde{Y}^2\right\rangle}\right)\\ &= -(1 - r_{XY}{}^2)\ln\left(4\pi^2 f^2\,(1 - r_{XY}{}^2)\left\langle \widetilde{X}^2\right\rangle\left\langle \widetilde{Y}^2\right\rangle\right) = -(1 - r_{XY}{}^2)\ln[(1 - r_{XY}{}^2)\,C],\end{aligned} \tag{10.54}$$

where the constant $C$ is defined by $C = 4\,\pi^2 f^2 < \widetilde{X}^2 > < \widetilde{Y}^2 >$. The meaning of this relation can be better seen by introducing the variables

$$x' = \frac{\hat{x} + \hat{y}}{\sqrt{2}}, \qquad\qquad y' = \frac{-\hat{x} + \hat{y}}{\sqrt{2}}. \tag{10.55}$$

The $(x', y')$-coordinate system is obtained by rotating the $(\hat{x}, \hat{y})$-coordinate system by an angle of 45°. The variables $\hat{x}$ and $\hat{y}$ are related to $x'$ and $y'$ by

$$\hat{x} = \frac{x' - y'}{\sqrt{2}}, \qquad\qquad \hat{y} = \frac{x' + y'}{\sqrt{2}}. \tag{10.56}$$

In terms of the latter expressions we can write the left-hand side of Eq. (10.54) as

$$\begin{aligned}\hat{x}^2 + \hat{y}^2 - 2r_{XY}\hat{x}\,\hat{y} &= \frac{1}{2}\left[(x' - y')^2 + (x' + y')^2 - 2r_{XY}(x' + y')(x' - y')\right]\\ &= \frac{1}{2}\left[2x'^2 + 2y'^2 - 2r_{XY}(x'^2 - y'^2)\right] = (1 - r_{XY})\,x'^2 + (1 + r_{XY})\,y'^2,\end{aligned} \tag{10.57}$$

such that Eq. (10.54) reads

$$(1 - r_{XY})\,x'^2 + (1 + r_{XY})\,y'^2 = -(1 - r_{XY}{}^2)\ln[(1 - r_{XY}{}^2)\,C]. \tag{10.58}$$

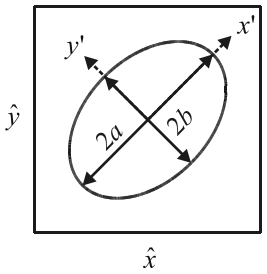



**Fig. 10.1.** Isolines of the joint normal PDF $f(x, y)$. The $(x', y')$-coordinate system is obtained by rotating the $(\hat{x}, \hat{y})$ -coordinate system by a 45° angle. The isoline $f(x, y) = f$ is an ellipse in the $(x', y')$-system. Here, $a$ is the semimajor axis, and $b$ is the semiminor axis.

The latter equation can be written

$$\frac{x'^2}{a^2} + \frac{y'^2}{b^2} = 1. \tag{10.59}$$

Here, the parameters $a$ and $b$ are given by

$$a = \sqrt{\frac{-(1-r_{XY}{}^2)\ln[(1-r_{XY}{}^2)C]}{1-r_{XY}}} = \sqrt{-\ln[(1-r_{XY}{}^2)C]^{1+r_{XY}}}, \tag{10.60a}$$

$$b = \sqrt{\frac{-(1-r_{XY}{}^2)\ln[(1-r_{XY}{}^2)C]}{1+r_{XY}}} = \sqrt{-\ln[(1-r_{XY}{}^2)C]^{1-r_{XY}}}. \tag{10.60b}$$

The relevance of Eq. (10.59) is that this relation represents an ellipse equation: see the illustration in Fig. 10.1. This ellipse equation involves two specific cases. For $r_{XY} = 0$ we find $a = b = [-\ln C]^{1/2}$, which means that the ellipse becomes a circle. The second case is given for $r_{XY} = \pm 1$: for $r_{XY} \to +1$ we have a line along the $x'$ axis, and for $r_{XY} \to -1$ we have a line along the $y'$ axis (see exercise 10.3.4).

**Filtered Joint PDF Calculation.** How can we numerically calculate the joint PDF $f(x, y)$ to test the suitability of model assumptions? In extension of the calculation of marginal PDFs we calculate the filtered joint PDF $f_\Delta(x, y)$ by

$$f_\Delta(x, y) = \frac{1}{\Delta x \Delta y}\frac{\Delta N_{xy}}{N}. \tag{10.61}$$

Here, $\Delta N_{xy}$ is the number of $(X, Y)$ realizations that are found in $x$ and $y$ intervals centered at $x$ and $y$. This means, $\Delta N_{xy}$ refers to the number of $(X, Y)$ realizations for which $X$ and $Y$ satisfy the conditions

$$x - \frac{\Delta x}{2} \le X \le x + \frac{\Delta x}{2} \qquad \text{and} \qquad y - \frac{\Delta y}{2} \le Y \le y + \frac{\Delta y}{2}. \tag{10.62}$$

It is relevant to note that $X$ and $Y$ are not any random values, but they represent a joint event (they are measured at the same time).

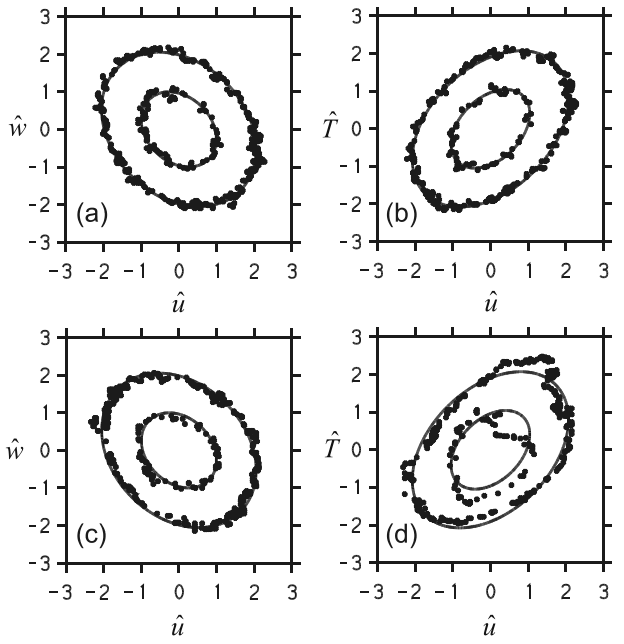



**Fig. 10.2.** Scatter plots of jointly normal and non-normal random variables. Here, $\hat{u}$, $\hat{w}$, and $\hat{T}$ are standardized velocity components $u$ and $w$ and temperature $T$ under neutral conditions, see the explanations given in Sect. 4.5. The correlation coefficient $r_{uw} = -0.26$ in (a) and (c), whereas $r_{uT} = 0.39$ in (b) and (d). The *scatter plots* in (a) and (b) show jointly normally distributed random variables, and the *scatter plots* in (c) and (d) are obtained from measurements described in Sect. 4.5. The *solid lines* represent isolines of jointly normally distributed variables. The outer and inner isolines correspond to $f \approx 0.02$ and $f \approx 0.1$, respectively.

**Scatter Plots of Normal Variables.** A good way to illustrate joint PDFs is to consider scatter plots of joint PDF isolines. Such scatter plots can be obtained by presenting all $(X, Y)$ positions for which the joint PDF $f_\Delta(x, y)$ has a certain value (or is found inside a certain interval). An example for this way of looking at the joint PDF is given in Figs. 10.2a–b. These figures show scatter plots of jointly normally distributed random numbers. These examples are set up according to Figs. 10.2c–d. Therefore, the variables are called $\hat{u}$, $\hat{w}$, and $\hat{T}$ ($\hat{u}$ and $\hat{w}$ refer to velocity components, and $\hat{T}$ refers to the temperature). The mean of these variables is zero and the variance is one, this means we consider standardized random variables. Figures 10.2a–b differ by their correlation coefficients, which have values that agree with the values in Figs. 10.2c–d. The joint PDF was calculated by using $\Delta x = \Delta y = 0.2$. As used for the Figs. 10.2c–d, a total number of 50,400

random numbers was considered. The outer isolines correspond exactly to the constant value $f = 40 / (N\,\Delta x\,\Delta y) \approx 0.02$, and the inner isolines correspond exactly to the constant value $f = 200 / (N\,\Delta x\,\Delta y) \approx 0.1$. Figures 10.2a–b show that the scatter plots obtained in this way agree very well with the isolines, which were calculated according to Eq. (10.59).

**Scatter Plots of Non-Normal Variables.** Such scatter plots of joint PDFs can be used to test the suitability of modeling measured data by a joint normal PDF. An illustration of this approach is given in Figs. 10.2c–d. These figures show joint PDFs of measured velocities and temperatures that were used in Sect. 4.5 to study marginal PDFs derived from measurements. We see here the joint $\hat{u} - \hat{w}$ PDF and the joint $\hat{u} - \hat{T}$ PDF for a neutral stratification. These joint PDFs have been calculated in the same way as the joint PDFs in Figs. 10.2a–b. The marginal PDFs of $u$ and $w$ shown in Fig. 4.17 reveal that both PDFs can be described very well by a normal PDF. Hence, the scatter plot in Fig. 10.2c agrees very well with the corresponding plot in Fig. 10.2a, which means that the joint $\hat{u} - \hat{w}$ PDF can be described very well by a joint normal PDF. As may be seen in Fig. 4.17, the marginal temperature PDF can be described only approximately by a normal PDF. Hence, the scatter plot in Fig. 10.2d also shows deviations to the corresponding joint normal PDF features given in Fig. 10.2b. Nevertheless, regarding the usual lack of alternatives it is still reasonable to describe the joint $\hat{u} - \hat{T}$ PDF by a joint normal PDF.

### *10.3.3 Application to Random Walk Modeling*

Let us consider the modeling of random walk (see Sect. 6.3) to illustrate the application of concepts introduced above. We consider a random variable (e.g., the position of any object) that is initially normally distributed. In each time step, the variable changes by the addition of a normally distributed contribution, which is independent of previous values of the random variable (it is worth emphasizing that the result to be obtained below can be extended to the case of jointly normally distributed variables that are correlated: see exercise 10.3.6). Hence, the random variable considered represents at every time a sum of independent normally distributed random numbers. The question related to this problem is to find the PDF of the variable considered at any time, this means the PDF of a sum of independent and normally distributed random numbers. This question, which requires the use of joint PDF concepts due to the need to consider simultaneously various random variables involved in the sum considered, will be addressed in the following.

**Sum of Two Variables.** First, let us consider the sum of two random variables with any statistical properties. In particular, we consider one variable $X_1$ with a

marginal PDF $f_1(x_1)$, and another variable $X_2$ with a marginal PDF $f_2(x_2)$. The joint PDF of both variables is given by $f_{12}(x_1, x_2)$. Our objective is to calculate the PDF $f(z)$ of the sum $Z = X_1 + X_2$. For doing this it is helpful to consider the distribution function $F(z)$, which enables the calculation of the PDF by means of $f(z) = dF / dz$. The distribution function $F(z)$ of the sum of two variables can be related to the joint PDF $f_{12}(x_1, x_2)$, which is considered to be known, by the relation

$$F(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x_1} f_{12}(x_1, x_2)\, dx_2\, dx_1. \tag{10.63}$$

Evidence for the validity of this relation can be obtained in the following way,

$$\begin{aligned} F(z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x_1} \left\langle \frac{d\theta(x_1 - X_1)}{dx_1} \frac{d\theta(x_2 - X_2)}{dx_2} \right\rangle dx_2\, dx_1 \\ &= \int_{-\infty}^{\infty} \left\langle \frac{d\theta(x_1 - X_1)}{dx_1} \left[\theta(z - x_1 - X_2) - \theta(-\infty - X_2)\right] \right\rangle dx_1 \\ &= \left\langle \int_{-\infty}^{z-X_2} \frac{d\theta(x_1 - X_1)}{dx_1} dx_1 \right\rangle \\ &= \left\langle \theta(z - X_2 - X_1) - \theta(-\infty - X_1) \right\rangle = \left\langle \theta(z - X_2 - X_1) \right\rangle. \end{aligned} \tag{10.64}$$

The first line applies the definition of $f_{12}(x_1, x_2)$. The integration with regard to $x_2$ is performed in the second line. The third line accounts for $\theta(-\infty - X_2) = 0$ and the fact that the integral is only nonzero if $x_1 \le z - X_2$. The brackets have to apply to all the integral now because the upper bound is a random number. The integration with regard to $x_1$ is performed in the fourth line, where $\theta(-\infty - X_1) = 0$ is used. The last expression represents $P(X_1 + X_2 \le z)$, which is the definition of $F(z)$. The corresponding PDF can be obtained by differentiating $F(z)$,

$$f(z) = \frac{dF(z)}{dz} = \int_{-\infty}^{\infty} f_{12}(x_1, z - x_1)\, dx_1. \tag{10.65}$$

For the case that $X_1$ and $X_2$ are independent, the last formula reads

$$f(z) = \int_{-\infty}^{\infty} f_1(x_1)\, f_2(z - x_1)\, dx_1. \tag{10.66}$$

**Sum of Two Independent Normal Variables.** Next, let us apply the definition (10.66) of $f(z)$ for the case that $X_1$ and $X_2$ are independent normally distributed random variables. The use of the normal PDF expression (4.72) results in

$$f(z) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left\{ -\frac{(x_1 - \mu_1)^2}{2\sigma_1^{\,2}} - \frac{(z - x_1 - \mu_2)^2}{2\sigma_2^{\,2}} \right\} dx_1. \tag{10.67}$$

Here, $\mu_1$ and $\sigma_1$ are the mean and standard deviation of $X_1$, and $\mu_2$ and $\sigma_2$ are the mean and standard deviation of $X_2$. By introducing $y = x_1 - \mu_1$ and replacing the integration over $x_1$ by an integration over $y$, $f(z)$ is given by

$$\begin{aligned} f(z) &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} \exp\left\{-\frac{y^2}{2\sigma_1{}^2} - \frac{(\hat{z}-y)^2}{2\sigma_2{}^2}\right\} dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma_1{}^2\sigma_2{}^2}\left[\sigma_2{}^2 y^2 + \sigma_1{}^2(\hat{z}-y)^2\right]\right\} dy, \end{aligned} \tag{10.68}$$

where the abbreviation $\hat{z} = z - \mu_1 - \mu_2$ is applied. We rewrite the bracket term to prepare the integration,

$$\begin{aligned} \sigma_2{}^2 y^2 + \sigma_1{}^2(\hat{z}-y)^2 &= (\sigma_1{}^2+\sigma_2{}^2)y^2 - 2\sigma_1{}^2\hat{z}\,y + \sigma_1{}^2\hat{z}^2 \\ &= (\sigma_1{}^2+\sigma_2{}^2)\left(y - \frac{\sigma_1{}^2\hat{z}}{\sigma_1{}^2+\sigma_2{}^2}\right)^2 + \sigma_1{}^2\hat{z}^2\left(1 - \frac{\sigma_1{}^2}{\sigma_1{}^2+\sigma_2{}^2}\right) \\ &= (\sigma_1{}^2+\sigma_2{}^2)\left(y - \frac{\sigma_1{}^2\hat{z}}{\sigma_1{}^2+\sigma_2{}^2}\right)^2 + \frac{\sigma_1{}^2\sigma_2{}^2\hat{z}^2}{\sigma_1{}^2+\sigma_2{}^2}. \end{aligned} \tag{10.69}$$

Therefore, $f(z)$ reads

$$f(z) = \frac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{\hat{z}^2}{2(\sigma_1{}^2+\sigma_2{}^2)}\right\}\int_{-\infty}^{\infty}\exp\left\{-\frac{\sigma_1{}^2+\sigma_2{}^2}{2\sigma_1{}^2\sigma_2{}^2}\left(y - \frac{\sigma_1{}^2\hat{z}}{\sigma_1{}^2+\sigma_2{}^2}\right)^2\right\} dy. \tag{10.70}$$

The integration can be performed by introducing the variable

$$s = \sqrt{\frac{\sigma_1{}^2+\sigma_2{}^2}{2\sigma_1{}^2\sigma_2{}^2}}\left(y - \frac{\sigma_1{}^2\hat{z}}{\sigma_1{}^2+\sigma_2{}^2}\right). \tag{10.71}$$

By replacing $y$ by $s$ in Eq. (10.70) we obtain

$$f(z) = \frac{1}{2\pi\sigma_1\sigma_2}\sqrt{\frac{2\sigma_1{}^2\sigma_2{}^2}{\sigma_1{}^2+\sigma_2{}^2}}\exp\left\{-\frac{\hat{z}^2}{2(\sigma_1{}^2+\sigma_2{}^2)}\right\}\int_{-\infty}^{\infty} e^{-s^2}\,ds. \tag{10.72}$$

The integral over $\exp(-s^2)$ is $\pi^{1/2}$ according to Eq. (4.70). Therefore, Eq. (10.72) reduces to

$$f(z) = \frac{1}{\sqrt{2\pi(\sigma_1{}^2+\sigma_2{}^2)}}\exp\left\{-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1{}^2+\sigma_2{}^2)}\right\}, \tag{10.73}$$

where $\hat{z} = z - \mu_1 - \mu_2$ is used. This expression shows that the PDF of the sum of two independent normally distributed variables is normal with mean $\mu_1 + \mu_2$ and

variance $\sigma_1 + \sigma_2$. This observation can be summarized by the conclusion

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \text{ and } X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2) \quad \Rightarrow \quad X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_1, \sigma_1^2 + \sigma_2^2). \tag{10.74}$$

$\mathcal{N}(\mu, \sigma^2)$ refers a normal PDF with mean $\mu$ and variance $\sigma^2$. The notation applied here means that $X_1$, $X_2$, and $X_1 + X_2$ are normally distributed with the means and variances specified by the corresponding $\mathcal{N}$.

**Sum of Independent Normal Variables.** The result (10.74) obtained for two independent normally distributed random variables $X_1$ and $X_2$ can be extended to the case of any number of independent normally distributed random variables. By considering the two numbers considered before as one number and adding another number, we find that the sum of three independent normally distributed variables is again normally distributed. Correspondingly, we can conclude ($i = 1, N$)

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \quad \Rightarrow \quad \sum_{i=1}^{N} X_i \sim \mathcal{N}\left( \sum_{i=1}^{N} \mu_i, \sum_{i=1}^{N} \sigma_i^2 \right). \tag{10.75}$$

Therefore, the PDF of the sum of $N$ independent normally distributed random variables represents a normal PDF. Its mean is given by the sum of all means, and its variance is given by the sum of all variances. The conclusion (10.75) obtained can be used for deriving a corresponding conclusion for the distribution of the mean value of $N$ independent normally distributed random variables. By replacing $X_i$ by $X_i / N$ we find that

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \quad \Rightarrow \quad \frac{1}{N} \sum_{i=1}^{N} X_i \sim \mathcal{N}\left( \frac{1}{N} \sum_{i=1}^{N} \mu_i, \frac{1}{N^2} \sum_{i=1}^{N} \sigma_i^2 \right). \tag{10.76}$$

Hence, the PDF of mean values is normally distributed, where the mean is given by the mean of all means involved and the variance is given by the mean of all variances involved divided by $N$. The latter results were applied in Sects. 6.2 and 6.3 for modeling a random walk (for determining the evolution of the position PDF in time).

## 10.4 The Fokker-Planck Equation

After considering the normal model for the joint PDF in the previous section let us consider now the modeling of the evolution of any PDF. This question will be addressed by generalizing the Fokker-Planck equation (8.21) for the PDF of one random variable to the case of any number of random variables. The question of how the Fokker-Planck equation is related to stochastic differential equations for the corresponding random variables will be discussed, too.

### *10.4.1 Definition of Multivariate Probability Density Functions*

The generalization of the Fokker-Planck equation (8.21) to an equation for the joint PDF of a vectorial stochastic process $\boldsymbol{X}(t) = \{X_1(t), X_2(t), \cdots, X_N(t)\}$ requires a relevant first step: the definition of a multivariate PDF $f(\boldsymbol{x}, t)$. The most efficient way of doing this is the use of theta and delta functions for several variables.

**Multivariate Theta and Delta Functions.** In Sect. 4.2.2 we introduced theta and delta functions of one variable. For a vectorial process $\boldsymbol{X}(t) = \{X_1(t), X_2(t), \cdots, X_N(t)\}$, the corresponding theta and delta functions are given by

$$\theta\big(\boldsymbol{x} - \boldsymbol{X}(t)\big) = \theta\big(x_1 - X_1(t)\big)\,\theta\big(x_2 - X_2(t)\big)\cdots\theta\big(x_N - X_N(t)\big), \tag{10.77a}$$

$$\delta\big(\boldsymbol{x} - \boldsymbol{X}(t)\big) = \delta\big(x_1 - X_1(t)\big)\,\delta\big(x_2 - X_2(t)\big)\cdots\delta\big(x_N - X_N(t)\big). \tag{10.77b}$$

Hence, multivariate theta and delta functions are products of all the theta and delta functions of single variables.

**Multivariate PDFs.** The last expression provides the basis for the definition of a multivariate PDF. By averaging (10.77b), the joint PDF $f(\boldsymbol{x}, t)$ can be defined by

$$f(\boldsymbol{x},t) = \big\langle \delta\big(\boldsymbol{x} - \boldsymbol{X}(t)\big)\big\rangle. \tag{10.78}$$

The brackets refer to the mean value defined by Eq. (4.1). The latter definition generalizes the definition (4.29) of the PDF of a single variable. In terms of the normalization property of delta functions we find that this definition satisfies the normalization condition for the joint PDF $f(\boldsymbol{x}, t)$,

$$\int f(\boldsymbol{x},t)\,d\boldsymbol{x} = \int \big\langle \delta\big(\boldsymbol{x} - \boldsymbol{X}(t)\big)\big\rangle\, d\boldsymbol{x} = \big\langle 1\big\rangle = 1. \tag{10.79}$$

Here, $d\boldsymbol{x} = dx_1\, dx_2 \cdots dx_N$ represents a multivariate differential given by the product of all differentials involved. Two-point PDFs can be defined correspondingly. For example, the two-point PDF $f(\boldsymbol{x}, t; \boldsymbol{x}', t')$ for having joint events $(\boldsymbol{x}, t)$ and $(\boldsymbol{x}', t')$ is defined by

$$f(\boldsymbol{x},t;\boldsymbol{x}',t') = \big\langle \delta\big(\boldsymbol{x} - \boldsymbol{X}(t)\big)\;\delta\big(\boldsymbol{x}' - \boldsymbol{X}(t')\big)\big\rangle. \tag{10.80}$$

The one-point PDF $f(\boldsymbol{x}, t)$ can be recovered from this definition,

$$f(\boldsymbol{x},t) = \int f(\boldsymbol{x},t;\boldsymbol{x}',t')\; d\boldsymbol{x}'. \tag{10.81}$$

The validity of this relation can be seen by using the definition (10.80) of the two-point PDF f$(\boldsymbol{x}, t; \boldsymbol{x}', t')$,

$$f(\boldsymbol{x},t) = \int \big\langle \delta\big(\boldsymbol{x} - \boldsymbol{X}(t)\big)\delta\big(\boldsymbol{x}' - \boldsymbol{X}(t')\big)\big\rangle\; d\boldsymbol{x}' = \big\langle \delta\big(\boldsymbol{x} - \boldsymbol{X}(t)\big)\big\rangle. \tag{10.82}$$

A PDF $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$ conditioned on $\boldsymbol{X}(t') = \boldsymbol{x}'$ can be defined in correspondence to the definition (8.37) for a single-variable PDF,

$$f(\boldsymbol{x},t\,|\,\boldsymbol{x}',t') = \frac{f(\boldsymbol{x},t;\boldsymbol{x}',t')}{f(\boldsymbol{x}',t')} = \frac{\left\langle \delta\big(\boldsymbol{x}-\boldsymbol{X}(t)\big)\,\delta\big(\boldsymbol{x}'-\boldsymbol{X}(t')\big)\right\rangle}{\left\langle \delta\big(\boldsymbol{x}'-\boldsymbol{X}(t')\big)\right\rangle} \\ = \left\langle \delta\big(\boldsymbol{x}-\boldsymbol{X}(t)\big)\,|\,\boldsymbol{X}(t')=\boldsymbol{x}'\right\rangle = \left\langle \delta\big(\boldsymbol{x}-\boldsymbol{X}(t)\big)\,|\,\boldsymbol{x}',t'\right\rangle. \tag{10.83}$$

In terms of this definition the one-point PDF $f(\boldsymbol{x}, t)$ can be written

$$f(\boldsymbol{x},t) = \int f(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')\; f(\boldsymbol{x}',t')\,d\boldsymbol{x}'. \tag{10.84}$$

This representation will be used in Sect. 10.4.3 for the derivation of solutions to the Fokker-Planck equation.

### *10.4.2 The Fokker-Planck Equation*

**Fokker-Planck Equation.** Let us consider an $N$-dimensional stochastic vector process $\boldsymbol{X}(t) = \{X_1(t), X_2(t), \cdots, X_N(t)\}$. This process is assumed to be Markovian and to have a continuous sample path. The extension of Eq. (8.21) to an equation for the joint PDF $f(\boldsymbol{x}, t)$ of the process $\boldsymbol{X}(t)$ reads

$$\frac{\partial f(\boldsymbol{x},t)}{\partial t} = -\frac{\partial D_i(\boldsymbol{x},t)\,f(\boldsymbol{x},t)}{\partial x_i} + \frac{\partial^2 D_{ij}(\boldsymbol{x},t)\,f(\boldsymbol{x},t)}{\partial x_i \partial x_j}. \tag{10.85}$$

Here, the sum convention is applied, this means the sum is taken over repeated subscripts. Equation (10.85) represents the Fokker-Planck equation for several variables (Fokker 1914, Planck 1917). Its coefficients $D_i$ and $D_{ij}$ are given by the vectorial generalizations of $D^{(1)}$ and $D^{(2)}$ given by Eq. (8.22),

$$D_i(\boldsymbol{x},t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left\langle X_i\big(t+\Delta t\big) - X_i\big(t\big)\,\big|\,\boldsymbol{x},t\right\rangle, \tag{10.86a}$$

$$D_{ij}(\boldsymbol{x},t) = \lim_{\Delta t \to 0} \frac{1}{2\Delta t}\left\langle \big[X_i\big(t+\Delta t\big) - X_i\big(t\big)\big]\big[X_j\big(t+\Delta t\big) - X_j\big(t\big)\big]\,\big|\,\boldsymbol{x},t\right\rangle. \tag{10.86b}$$

The conditional means refer to the condition $\boldsymbol{X}(t) = \boldsymbol{x}$. Equation (10.85) has the structure of a diffusion equation. The coefficient $D_i$ represents a drift coefficient and $D_{ij}$ is a diffusion coefficient. The coefficient $D_{ij}$ has two relevant properties, which are a consequence of its definition (10.86b). The first property is that $D_{ij}$ is symmetric, this means $D_{ij} = D_{ji}$. The second property is that $D_{ij}$ is positive semi-definite, this means $D_{ij}$ is non-negative definite. This property of $D_{ij}$ can be shown by multiplying the definition (10.86b) with arbitrary real nonvanishing vectors $c_i$

and $c_j$, which results in

$$\begin{aligned} D_{ij} c_i c_j &= \lim_{\Delta t \to 0} \frac{1}{2\Delta t} \left\langle \left[ X_i(t+\Delta t) - X_i(t) \right] c_i \left[ X_j(t+\Delta t) - X_j(t) \right] c_j \,\middle|\, \boldsymbol{x}, t \right\rangle \\ &= \lim_{\Delta t \to 0} \frac{1}{2\Delta t} \left\langle \left( \left[ X_i(t+\Delta t) - X_i(t) \right] c_i \right)^2 \,\middle|\, \boldsymbol{x}, t \right\rangle \geq 0. \end{aligned} \tag{10.87}$$

Usually, it is assumed that $D_{ij}$ is positive definite, this means

$$D_{ij} c_i c_j > 0. \tag{10.88}$$

The inverse matrix of $D_{ij}$ will exist for this case (Pope 2000), which is relevant to solutions of the Fokker-Planck equation. A positive definite matrix has positive eigenvalues, as may be seen in the following way (Ortega 1987): Suppose that $\lambda$ is an eigenvalue of the matrix $D_{ij}$ and $c_j$ is a corresponding real nonvanishing eigenvector, this means $D_{ij} c_j = \lambda c_i$. Multiplication of both sides with $c_i$ provides $D_{ij} c_i c_j = \lambda c_i c_i$. Therefore, we find $\lambda = D_{ij} c_i c_j / (c_i c_i) > 0$. The existence of positive eigenvalues is a necessary and sufficient condition for a positive definite matrix $D_{ij}$. For positive eigenvalues we find that the three principal invariants of $D_{ij}$ (one of the invariants is the determinant det($D$) of $D_{ij}$) have to be positive: see exercise 10.4.1. On the other hand, three positive principal invariants imply that the matrix $D_{ij}$ has to be positive definite.

**Consistency Constraint.** The consistency of the Fokker-Planck equation can be proven by integrating Eq. (10.85) over the sample space $\boldsymbol{x}$,

$$\int \frac{\partial f(\boldsymbol{x},t)}{\partial t} d\boldsymbol{x} = -\int \frac{\partial D_i(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_i} d\boldsymbol{x} + \int \frac{\partial^2 D_{ij}(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_i \partial x_j} d\boldsymbol{x}. \tag{10.89}$$

The left-hand side of Eq. (10.89) vanishes: we can write the time derivative in front of the integral, and $f(\boldsymbol{x}, t)$ is normalized to one. The terms on the right-hand side can be treated by invoking the Divergence Theorem. This theorem states the following (Stewart 2006): Let $E$ be a simple solid region and let $S$ be the boundary surface of $E$, given with positive (outward) orientation. Let $\boldsymbol{L}$ be a vector field whose component functions have continuous partial derivatives on an open region that contains $E$. Then

$$\int_S \boldsymbol{L} \cdot d\boldsymbol{S} = \int_E \frac{\partial L_i}{\partial x_i} d\boldsymbol{x}. \tag{10.90}$$

The Divergence Theorem can be applied to the right-hand side of Eq. (10.89) by setting $L_i = D_i f$ and $L_i = \partial(D_{ij} f) / \partial x_j$, respectively. By considering an infinite domain, the integrals on the right-hand side of Eq. (10.89) will vanish if $L_i$ is zero at the surface. Therefore, the consistency of the Fokker-Planck equation (10.85) requires the assumption that the PDF $f(\boldsymbol{x}, t)$ and its derivatives vanish for $|\boldsymbol{x}| \to \infty$.

**Mean Equations.** By multiplying the Fokker-Planck equation (10.85) with $x_k$ and integration over the sample space we obtain

$$\int x_k \frac{\partial f(\boldsymbol{x},t)}{\partial t} d\boldsymbol{x} = -\int x_x \frac{\partial D_i(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_i} d\boldsymbol{x} + \int x_k \frac{\partial^2 D_{ij}(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_i \partial x_j} d\boldsymbol{x}. \quad (10.91)$$

To enable the rewriting of the right-hand side and to prepare the use of Eq. (10.90) we write this equation as

$$\begin{aligned} \frac{\partial}{\partial t}\int x_k f(\boldsymbol{x},t) d\boldsymbol{x} = &-\int \frac{\partial x_k D_i(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_i} d\boldsymbol{x} + \int \frac{\partial x_k}{\partial x_i} D_i(\boldsymbol{x},t) f(\boldsymbol{x},t) d\boldsymbol{x} \\ &+ \int \frac{\partial}{\partial x_i}\left[ x_k \frac{\partial D_{ij}(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_j} \right] d\boldsymbol{x} - \int \frac{\partial x_k}{\partial x_i} \frac{\partial D_{ij}(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_j} d\boldsymbol{x}. \end{aligned} \quad (10.92)$$

The integral on the left-hand side is equal to the mean $\langle X_k \rangle$. The validity of the right-hand side can be seen by distributing the derivatives by $x_i$ involved in the first and third terms. For $\partial x_k / \partial x_i$ we find $\partial x_k / \partial x_i = \delta_{ki}$. Here, $\delta_{ik}$ is the Kronecker symbol, which has the properties $\delta_{ik} = 1$ for $i = k$ and $\delta_{ik} = 0$ for $i \neq k$. By accounting for $\partial x_k / \partial x_i = \delta_{ki}$, three of the four terms on the right-hand side of Eq. (10.92) can be written as integrals over the surface $S$ according to Eq. (10.90). We assume that the corresponding terms disappear for $|\boldsymbol{x}| \to \infty$ so that Eq. (10.92) reads

$$\frac{\partial \langle X_k \rangle}{\partial t} = \delta_{ki} \int D_i(\boldsymbol{x},t) f(\boldsymbol{x},t) d\boldsymbol{x} = \int D_k(\boldsymbol{x},t) f(\boldsymbol{x},t) d\boldsymbol{x}. \quad (10.93)$$

The last expression is implied by the fact that $\delta_{ki}$ is only nonzero for $k = i$. The right-hand side represents the mean value $\langle D_k \rangle$. Hence we find

$$\frac{d\langle X_k \rangle}{dt} = \langle D_k \rangle. \quad (10.94)$$

The partial derivative by $t$ can be replaced here by the regular derivative because $\langle X_k \rangle$ and $\langle D_k \rangle$ are only functions of $t$. Hence, $\langle D_k \rangle$ determines the transport of means $\langle X_k \rangle$. For that reason $D_k$ is called a drift coefficient.

**Variance Equations.** The variance equations can be obtained by multiplying the Fokker-Planck equation (10.85) with $x_k x_n$ and integrating over $\boldsymbol{x}$,

$$\int x_k x_n \frac{\partial f(\boldsymbol{x},t)}{\partial t} d\boldsymbol{x} = -\int x_x x_n \frac{\partial D_i(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_i} d\boldsymbol{x} + \int x_k x_n \frac{\partial^2 D_{ij}(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_i \partial x_j} d\boldsymbol{x}. \quad (10.95)$$

This equation can be also written

$$\frac{\partial \langle X_k X_n \rangle}{\partial t} = I_1 + I_2. \quad (10.96)$$

We wrote here the partial derivative by $t$ in front of the integral and applied the definition of $<X_k X_n>$. The symbols $I_1$ and $I_2$ refer to the first and second integral on the right-hand side of Eq. (10.95), respectively. To calculate $I_1$ we write

$$\begin{aligned} I_1 &= -\int \frac{\partial x_k x_n D_i(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_i} d\boldsymbol{x} + \int \frac{\partial x_k x_n}{\partial x_i} D_i(\boldsymbol{x},t) f(\boldsymbol{x},t) d\boldsymbol{x} \\ &= \int (\delta_{ki} x_n + \delta_{ni} x_k) D_i(\boldsymbol{x},t) f(\boldsymbol{x},t) d\boldsymbol{x} = \left\langle X_n D_k \right\rangle + \left\langle X_k D_n \right\rangle. \end{aligned} \tag{10.97}$$

The first rewriting identifies an integral over a derivative (the first term on the right-hand side), which disappears. The next rewriting accounts for $\partial(x_k\, x_n) / \partial x_i = \delta_{ki}\, x_n + \delta_{ni}\, x_k$. The definition of means and the property of the Kronecker symbol $\delta_{ki}$ to be nonzero only for $k = i$ are used for obtaining the final expression. The integral $I_2$ can be calculated correspondingly,

$$\begin{aligned} I_2 &= \int \frac{\partial}{\partial x_i} \left[ x_k x_n \frac{\partial D_{ij}(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_j} \right] d\boldsymbol{x} - \int \frac{\partial x_k x_n}{\partial x_i} \frac{\partial D_{ij}(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_j} d\boldsymbol{x} \\ &= -\int (\delta_{ki} x_n + \delta_{ni} x_k) \frac{\partial D_{ij}(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_j} d\boldsymbol{x} \\ &= -\int \frac{\partial (\delta_{ki} x_n + \delta_{ni} x_k) D_{ij}(\boldsymbol{x},t) f(\boldsymbol{x},t)}{\partial x_j} d\boldsymbol{x} + \int \frac{\partial (\delta_{ki} x_n + \delta_{ni} x_k)}{\partial x_j} D_{ij}(\boldsymbol{x},t) f(\boldsymbol{x},t) d\boldsymbol{x} \\ &= \int (\delta_{ki}\delta_{nj} + \delta_{ni}\delta_{kj}) D_{ij}(\boldsymbol{x},t) f(\boldsymbol{x},t) d\boldsymbol{x} = \left\langle D_{kn} \right\rangle + \left\langle D_{nk} \right\rangle = 2\left\langle D_{nk} \right\rangle. \end{aligned} \tag{10.98}$$

The last expression applies the symmetry of $D_{nk}$. The combination of Eq. (10.96) with these expressions for $I_1$ and $I_2$ leads then to the variance equation

$$\frac{d\left\langle X_k X_n \right\rangle}{dt} = \left\langle X_n D_k \right\rangle + \left\langle X_k D_n \right\rangle + 2\left\langle D_{nk} \right\rangle, \tag{10.99}$$

where the partial derivative by $t$ was replaced by the regular derivative. Instead of considering equations for second-order moments, it is more convenient to derive equations for the variance

$$\left\langle \tilde{X}_k \tilde{X}_n \right\rangle = \left\langle \left( X_k - \left\langle X_k \right\rangle \right) \left( X_n - \left\langle X_n \right\rangle \right) \right\rangle = \left\langle X_k X_n \right\rangle - \left\langle X_k \right\rangle \left\langle X_n \right\rangle. \tag{10.100}$$

By differentiating this variance expression we obtain

$$\frac{d\left\langle \tilde{X}_k \tilde{X}_n \right\rangle}{dt} = \frac{d\left\langle X_k X_n \right\rangle}{dt} - \frac{d\left\langle X_k \right\rangle \left\langle X_n \right\rangle}{dt} = \frac{d\left\langle X_k X_n \right\rangle}{dt} - \frac{d\left\langle X_k \right\rangle}{dt} \left\langle X_n \right\rangle - \frac{d\left\langle X_n \right\rangle}{dt} \left\langle X_k \right\rangle. \tag{10.101}$$

The use of Eqs. (10.94) and (10.99) implies then the following variance equations,

$$\begin{aligned}\frac{d\left\langle \widetilde{X}_k \widetilde{X}_n \right\rangle}{dt} &= \left\langle X_n D_k \right\rangle + \left\langle X_k D_n \right\rangle + 2\left\langle D_{nk} \right\rangle - \left\langle D_k \right\rangle\left\langle X_n \right\rangle - \left\langle D_n \right\rangle\left\langle X_k \right\rangle \\ &= \left\langle \widetilde{X}_n \widetilde{D}_k \right\rangle + \left\langle \widetilde{X}_k \widetilde{D}_n \right\rangle + 2\left\langle D_{nk} \right\rangle,\end{aligned} \tag{10.102}$$

where the variance expression (10.100) is used for obtaining the last expression. The variance of one component is given by setting $k = n$. We have $\langle D_{kk} \rangle \geq 0$ as a consequence of the definition (10.86b) of $D_{kn}$. Hence, variances are produced by $\langle D_{kn} \rangle$: a nonzero $D_{kn}$ causes a diffusion process (the width of the PDF increases). For that reason $D_{kn}$ is called a diffusion coefficient. An equilibrium state may be reached asymptotically if the first two terms on the right-hand side of Eq. (10.102) appear with a negative sign, i.e., if these terms model a dissipation of variance.

**Correlations.** The Fokker-Planck equation (10.85) can be used to calculate the correlation between $X_i(t)$ and $X_j(t')$. We assume that $t \leq t' = t + r$, where $r$ is any non-negative time. By following the derivation of the corresponding correlation (8.33) for the case of one variable (see exercise 8.3.1), the correlation of $X_i(t)$ and $X_j(t + r)$ is found to be determined by the equation (see also Eq. (10.124))

$$\frac{d\left\langle \widetilde{X}_i(t) \widetilde{X}_j(t+r) \right\rangle}{dr} = \left\langle \widetilde{X}_i(t)\, \widetilde{D}_j\left(\boldsymbol{X}(t+r), t+r\right) \right\rangle. \tag{10.103}$$

Thus, the correlation is unaffected by the diffusion coefficient $D_{ij}$, i.e., correlations are not produced, but they relax according to the model provided by $D_j$.

### *10.4.3 A Solution to the Fokker-Planck Equation*

**Equation Considered.** Let us illustrate the application of the Fokker-Planck equation (10.85) and demonstrate characteristic solution properties by considering an example that enables the derivation of an analytical solution. The equation considered is a vectorial generalization of Eq. (8.34) for a single variable,

$$\frac{\partial f(\boldsymbol{x},t)}{\partial t} = -\frac{\partial}{\partial x_i}\left[G_i(t) + G_{ik}(t)\left(x_k - \left\langle X_k \right\rangle\right)\right] f(\boldsymbol{x},t) + \frac{\partial^2 D_{ij}(t)\, f(\boldsymbol{x},t)}{\partial x_i \partial x_j}. \tag{10.104}$$

The drift coefficient $D_i$ is a linear function of the variables $\boldsymbol{x}$, which may be seen as first-order Taylor series of $D_i$. The inclusion of $\langle X_k \rangle$ in Eq. (10.104) defines $G_{ik}$ as the coefficient that controls the intensity of fluctuations about the mean $\langle X_k \rangle$. This linear model for $D_i$ is well suited for the characterization of near-equilibrium processes. The diffusion coefficient $D_{ij}$ is assumed to be only a function of time,

which is a convenient choice with regard to many applications. $D_{ij}$ is assumed to be positive definite. Equation (10.104) will be combined with the assumption of natural boundary conditions, this means $f(\boldsymbol{x}, t) \to 0$ as $|\boldsymbol{x}| \to \infty$.

**Solution Approach.** Solutions $f(\boldsymbol{x}, t)$ to the Fokker-Planck equation (10.104) will depend on the initial PDF $f(\boldsymbol{x}', t')$, which has to be provided. The influence of the initial PDF can be treated separately from the solution of the Fokker-Planck equation, which is very helpful for using solutions for a variety of initial PDFs. This can be achieved by using Eq. (10.84), which represents the one-point PDF $f(\boldsymbol{x}, t)$ in terms of the PDF conditioned on the initial condition $\boldsymbol{X}(t') = \boldsymbol{x}'$,

$$f(\boldsymbol{x},t) = \int f(\boldsymbol{x},t \,|\, \boldsymbol{x}',t')\; f(\boldsymbol{x}',t')\, d\boldsymbol{x}'. \tag{10.105}$$

The idea of this approach is to calculate a general expression for the conditional PDF $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$ independent of the initial PDF $f(\boldsymbol{x}', t')$, and to calculate then the PDF $f(\boldsymbol{x}, t)$ by integration of Eq. (10.105). But how can we calculate $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$? In terms of Eq. (10.105), the Fokker-Planck equation (10.104) can be written

$$\begin{aligned} 0 = \int \Bigg\{ & \frac{\partial f(\boldsymbol{x},t \,|\, \boldsymbol{x}',t')}{\partial t} + \frac{\partial}{\partial x_i}\Big[G_i(t) + G_{ik}(t)\big(x_k - \langle X_k \rangle\big)\Big] f(\boldsymbol{x},t \,|\, \boldsymbol{x}',t') \\ & - \frac{\partial^2 D_{ij}(t)\, f(\boldsymbol{x},t \,|\, \boldsymbol{x}',t')}{\partial x_i \partial x_j} \Bigg\} f(\boldsymbol{x}',t')\; d\boldsymbol{x}'. \end{aligned} \tag{10.106}$$

Hence, the conditional PDF $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$ has to satisfy, too, the Fokker-Planck equation (10.104), i.e., the conditional PDF $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$ has to satisfy the equation

$$\begin{aligned} \frac{\partial f(\boldsymbol{x},t \,|\, \boldsymbol{x}',t')}{\partial t} = & - \frac{\partial\Big[G_i(t) + G_{ik}(t)\big(x_k - \langle X_k \rangle\big)\Big] f(\boldsymbol{x},t \,|\, \boldsymbol{x}',t')}{\partial x_i} \\ & + \frac{\partial^2 D_{ij}(t)\, f(\boldsymbol{x},t \,|\, \boldsymbol{x}',t')}{\partial x_i \partial x_j}. \end{aligned} \tag{10.107}$$

Eq. (10.83) provides the initial condition for the conditional PDF $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$,

$$\begin{aligned} f(\boldsymbol{x},t' | \boldsymbol{x}',t') &= \frac{\big\langle \delta\big(\boldsymbol{x} - \boldsymbol{X}(t')\big)\delta\big(\boldsymbol{x}' - \boldsymbol{X}(t')\big)\big\rangle}{\big\langle \delta\big(\boldsymbol{x}' - \boldsymbol{X}(t')\big)\big\rangle} = \frac{\big\langle \delta(\boldsymbol{x} - \boldsymbol{x}')\,\delta\big(\boldsymbol{x}' - \boldsymbol{X}(t')\big)\big\rangle}{\big\langle \delta\big(\boldsymbol{x}' - \boldsymbol{X}(t')\big)\big\rangle} \\ &= \delta(\boldsymbol{x} - \boldsymbol{x}'), \end{aligned} \tag{10.108}$$

where the sifting property of delta functions is used.

**Conditional PDF Calculation.** The conditional PDF is a normal PDF for the single-variable case (see Sect. 8.3.2). Therefore, we may assume that $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$ also is given by a normal PDF (an $N$-dimensional normal PDF for our case),

$$f(\boldsymbol{x},t \,|\, \boldsymbol{x}',t') = \frac{1}{(2\pi)^{N/2}\sqrt{\det(\beta)}} \exp\left\{ -\frac{1}{2}\beta^{-1}{}_{ij}\big(x_i - \alpha_i\big)\big(x_j - \alpha_j\big) \right\}. \tag{10.109}$$

Here, $\alpha_i$ are the mean values and $\beta_{ij}$ represent the elements of the variance matrix, which is positive definite and symmetric ($\beta_{ij} = \beta_{ji}$). Therefore, the inverse matrix ${\beta^{-1}}_{ij}$ does exist, and it is symmetric, i.e., ${\beta^{-1}}_{ij} = {\beta^{-1}}_{ji}$. Another view of looking at the assumption (10.109) is the following one: we ask under which conditions it is possible to have a normal PDF as solution of a Fokker-Planck equation. To prove the suitability of the assumption (10.109) we have to show that Eq. (10.109) can satisfy Eq. (10.107). This fact is proven in terms of exercise 10.4.4. It is found that the model parameters $\alpha_i$ and $\beta_{ij}$ have to satisfy the equations

$$\frac{d\alpha_i}{dt} = G_i + G_{ik}\left(\alpha_k - \langle X_k \rangle\right), \tag{10.110a}$$

$$\frac{d\beta_{ij}}{dt} = G_{ik}\beta_{kj} + G_{jk}\beta_{ki} + 2D_{ij}. \tag{10.110b}$$

These initial conditions for $\alpha_i$ and $\beta_{ij}$ are given by

$$\alpha_i(t') = x'_i, \tag{10.111a}$$

$$\beta_{ij}(t') = 0. \tag{10.111b}$$

**Means and Variances Implied by the Fokker-Planck Equation.** Next, let us have a look at the means and variances of $f(\boldsymbol{x}, t)$, which are implied by the Fokker-Planck equation (10.104). The simplest way to obtain these equations is to specify the general Eqs. (10.94) and (10.102), which are valid for every Fokker-Planck equation,

$$\frac{d\langle X_k \rangle}{dt} = G_k, \tag{10.112a}$$

$$\frac{d\langle \tilde{X}_k \tilde{X}_n \rangle}{dt} = G_{km}\langle \tilde{X}_m \tilde{X}_n \rangle + G_{nm}\langle \tilde{X}_m \tilde{X}_k \rangle + 2D_{nk}. \tag{10.112b}$$

Equations (10.112) are similar to Eqs. (10.110) for the parameters $\alpha_k$ and $\beta_{kn}$ of the conditional PDF. To see the difference, we apply Eqs. (10.110) and (10.112) for deriving the following equations

$$\frac{d}{dt}\left(\alpha_{\mathrm{k}} - \langle X_k \rangle\right) = G_{kn}\left(\alpha_n - \langle X_n \rangle\right), \tag{10.113a}$$

$$\frac{d}{dt}\left(\beta_{\mathrm{kn}} - \langle \tilde{X}_k \tilde{X}_n \rangle\right) = G_{km}\left(\beta_{\mathrm{mn}} - \langle \tilde{X}_m \tilde{X}_n \rangle\right) + G_{nm}\left(\beta_{\mathrm{mk}} - \langle \tilde{X}_m \tilde{X}_k \rangle\right). \tag{10.113b}$$

The coefficient $G_{km}$ is usually provided with a negative sign to model a relaxation of fluctuations. For this case, $\alpha_k$ and $\beta_{kn}$ relax to the means and variances of $f(\mathbf{x}, t)$:

the stationary values of $\alpha_k$ and $\beta_{kn}$, for which the left-hand sides of Eqs. (10.113) are zero, are given by

$$\alpha_{\mathrm{k}} = \left\langle X_k \right\rangle, \tag{10.114a}$$

$$\beta_{\mathrm{kn}} = \left\langle \widetilde{X}_k \widetilde{X}_n \right\rangle. \tag{10.114b}$$

For this case, the conditional PDF $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$ is independent of $\boldsymbol{x}'$ because its parameters are independent of $\boldsymbol{x}'$. Equation (10.105) reveals that the PDF $f(\boldsymbol{x}, t)$ is then equal to the conditional PDF $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$. Therefore, the unconditional PDF $f(\boldsymbol{x}, t)$, which may have any shape initially, does relax (independent of the initial conditions) asymptotically to a normal PDF.

### *10.4.4 Stochastic Differential Equations*

**Stochastic Differential Equations.** In analogy to the discussion of the relationship between the Fokker-Planck equation (8.21) and the stochastic differential equation (8.55) for a single variable, let us consider now the corresponding relationship for several variables. For the case of an $N$-dimensional stochastic process $\boldsymbol{X}(t) = \{X_1(t), X_2(t), \cdots, X_N(t)\}$ we generalize the Markovian stochastic equation (8.55) by the equation

$$\frac{dX_i}{dt}(t) = a_i(\boldsymbol{X}, t) + b_{ik}(\boldsymbol{X}, t)\frac{dW_k}{dt}(t). \tag{10.115}$$

Here, the coefficients $a_i(\boldsymbol{X}(t), t)$ and $b_{ik}(\boldsymbol{X}(t), t)$ are any deterministic functions of $\boldsymbol{X}(t)$ and $t$. The normally distributed vectorial process $dW_k / dt$ is characterized by

$$\left\langle \frac{dW_k}{dt}(t) \right\rangle = 0, \tag{10.116a}$$

$$\left\langle \frac{dW_k}{dt}(t)\frac{dW_n}{dt}(t') \right\rangle = \delta_{kn}\,\delta(t - t'). \tag{10.116b}$$

Relation (10.116a) corresponds to Eq. (8.57). Relation (10.116b) corresponds to Eq. (8.60) for $k = n$. For $k \neq n$ this relation means that $dW_k / dt$ is uncorrelated to $dW_n / dt$. The process $dW_k / dt$ is assumed to be independent of $\boldsymbol{X}(t_0)$. Due to the fact that the change of the stochastic process $\boldsymbol{X}(t)$ is fully determined by $a_i(\boldsymbol{X}(t), t)$, $b_{ik}(\boldsymbol{X}(t), t)$, and $dW_k / dt$, we find that the equation system (10.115) describes the evolution of $\boldsymbol{X}(t)$ as a Markov process: the future of the statistical properties of $\boldsymbol{X}(t)$ is fully determined by the present state.

**Stochastic Difference Equations.** The representation of the stochastic differential equation (10.115) as a stochastic difference equation is relevant, e.g., to the numerical solution of Eq. (10.115) and regarding the derivation of the relationship to the Fokker-Planck equation (10.85): see the discussion in the next paragraph. To address this question we integrate Eq. (10.115) from $t$ to $t + \Delta t$,

$$X_i(t+\Delta t) - X_i(t) = \int_t^{t+\Delta t} a_i\big(\boldsymbol{X}(s), s\big)\, ds + \int_t^{t+\Delta t} b_{ik}\big(\boldsymbol{X}(s), s\big) \frac{dW_k}{ds}(s)\, ds, \tag{10.117}$$

where $\Delta t$ is a sufficiently small time interval. As for the single-variable case we use the Itô definition of stochastic integration, i.e., we approximate $a_i(\boldsymbol{X}(s), s)$ and $b_{ik}(\boldsymbol{X}(s), s)$ by their values at the lower bound $t$. Then, Eq. (10.117) can be written

$$X_i(t+\Delta t) - X_i(t) = a_i\big(\boldsymbol{X}(t), t\big)\Delta t + b_{ik}\big(\boldsymbol{X}(t), t\big)\, \Delta W_k(t), \tag{10.118}$$

where $\Delta W_k(t)$ is defined by

$$\Delta W_k(t) = \int_t^{t+\Delta t} \frac{dW_k}{ds}(s)\, ds = W_k(t+\Delta t) - W_k(t). \tag{10.119}$$

The properties of $\Delta W_k(t)$ can be derived in terms of Eq. (10.116),

$$\big\langle \Delta W_k(t) \big\rangle = \int_t^{t+\Delta t} \left\langle \frac{dW_k}{ds}(s) \right\rangle ds = 0. \tag{10.120a}$$

$$\begin{aligned} \big\langle \Delta W_k(t)\, \Delta W_n(t') \big\rangle &= \int_t^{t+\Delta t} \int_{t'}^{t'+\Delta t} \left\langle \frac{dW_k}{ds}(s) \frac{dW_n}{ds'}(s') \right\rangle ds'\, ds \\ &= \delta_{kn} \int_t^{t+\Delta t} \int_{t'}^{t'+\Delta t} \delta(s'-s)\, ds'\, ds = \delta_{kn} \int_t^{t+\Delta t} \int_{t'}^{t'+\Delta t} \frac{d\theta(s'-s)}{ds'}\, ds'\, ds \\ &= \delta_{kn} \int_t^{t+\Delta t} \big[\theta(t'+\Delta t - s) - \theta(t'-s)\big]\, ds = \delta_{kn}\, \Delta t \begin{Bmatrix} 1 & \text{if } t = t' \\ 0 & \text{if } t \neq t' \end{Bmatrix}. \end{aligned} \tag{10.120b}$$

Here, $t'$ changes by $\Delta t$ as does $t$, this means $t' = t + k\, \Delta t$, where $k = 0, \pm 1, \pm 2, \ldots$. Thus, the integral in the last line is only nonzero and equal to $\Delta t$ if $t' = t$, which explains the final result of Eq. (10.120b).

**Relationship to Fokker-Planck Equation.** The question about the relationship to the Fokker-Planck equation (10.85) can be addressed by the calculation of the first two coefficients of the Kramers-Moyal equation (which has to be written for the case of many variables). According to Eq. (10.86), these coefficients become

$$D_i(\boldsymbol{x}, t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \big\langle X_i(t+\Delta t) - X_i(t) \big| \boldsymbol{x}, t \big\rangle, \tag{10.121a}$$

$$D_{ij}(\boldsymbol{x}, t) = \lim_{\Delta t \to 0} \frac{1}{2\Delta t} \big\langle \big[X_i(t+\Delta t) - X_i(t)\big]\big[X_j(t+\Delta t) - X_j(t)\big] \big| \boldsymbol{x}, t \big\rangle. \tag{10.121b}$$

The use of Eq. (10.118) in these expressions leads to

$$D_i(\boldsymbol{x},t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left\langle a_i(\boldsymbol{X}(t),t)\Delta t + b_{ik}(\boldsymbol{X}(t),t)\,\Delta W_k(t) \,\middle|\, \boldsymbol{x},t \right\rangle = \mathrm{a_i}(\boldsymbol{x},\mathrm{t}), \quad (10.122a)$$

$$\begin{aligned} D_{ij}(\boldsymbol{x},t) = \lim_{\Delta t \to 0} \frac{1}{2\Delta t} \langle & \left[ a_i(\boldsymbol{X}(t),t)\Delta t + b_{ik}(\boldsymbol{X}(t),t)\,\Delta W_k(t) \right] \\ & \left[ a_j(\boldsymbol{X}(t),t)\Delta t + b_{jn}(\boldsymbol{X}(t),t)\,\Delta W_n(t) \right] \Big| \boldsymbol{x},t \Big\rangle \qquad (10.122b) \\ = \frac{1}{2} b_{ik}(\boldsymbol{x},t)\, b_{jn}(\boldsymbol{x},t)\, \delta_{kn} &= \frac{1}{2} b_{ik}(\boldsymbol{x},t)\, b_{jk}(\boldsymbol{x},t), \end{aligned}$$

where the properties (10.120) of $\Delta W_k$ are used. The corresponding calculation of higher-order coefficients of the multivariate Kramers-Moyal equation leads to the same conclusion as obtained for the single-variable case: all these coefficients are zero because they are of higher order in $\Delta t$. Thus, the stochastic Eq. (10.115) does imply uniquely a Fokker-Planck equation that determines the PDF evolution. However, a Fokker-Planck equation does not fully determine a stochastic differential equation in general. For $N$ variables, Eq. (10.122b) provides $N(N+1)/2$ equations for $N^2$ elements of $b_{ij}$ (e.g., for $N = 6$ there are only 21 equations for 36 elements of $b_{ij}$). Therefore, the coefficients of the stochastic Eq. (10.115) are only uniquely determined by the Fokker-Planck coefficients $D_i$ and $D_{ij}$ if $b_{ij}$ is assumed to be symmetric so that only $N(N+1)/2$ elements of $b_{ij}$ have to be determined.

**Correlations.** It was shown in the previous paragraph that the stochastic differential equation (10.115) is consistent with the Fokker-Planck equation (10.85) with regard to the one-point statistics (i.e., the PDF, means and variances), but the corresponding consistency regarding the correlation dynamics is not demonstrated in this way. To address this question we use the stochastic differential equation for deriving an equation for the correlation function of $X_i(t)$ and $X_j(t+r)$, where $r$ is any non-negative time. For doing this we consider

$$\begin{aligned} \frac{d\left\langle \widetilde{X}_i(t) X_j(t+r) \right\rangle}{dr} &= \left\langle \widetilde{X}_i(t) \frac{dX_j(t+r)}{dt} \right\rangle \\ &= \left\langle \widetilde{X}_i(t)\, a_j(\boldsymbol{X}(t+r),t+r) \right\rangle + \left\langle \widetilde{X}_i(t)\, b_{jk}(\boldsymbol{X}(t+r),t+r) \frac{dW_k}{dt}(t+r) \right\rangle. \end{aligned} \quad (10.123)$$

The last equation arises from the use of Eq. (10.115). The noise term $dW_k/dt(t+r)$ is independent of $\boldsymbol{X}(t)$ and $\boldsymbol{X}(t+r)$ because only noise at times before $t$ and $t+r$ can affect $\boldsymbol{X}(t)$ and $\boldsymbol{X}(t+r)$, respectively. Thus, the last term is zero and we obtain

$$\frac{d\left\langle \widetilde{X}_i(t) \widetilde{X}_j(t+r) \right\rangle}{dr} = \left\langle \widetilde{X}_i(t)\, \widetilde{a}_j(\boldsymbol{X}(t+r),t+r) \right\rangle. \quad (10.124)$$

Here, $X_j$ and $a_j$ are replaced by the corresponding fluctuations because the means of $X_j$ and $a_j$ do not affect the result. The result obtained generalizes Eq. (8.72) to the case of several variables, and it recovers Eq. (10.103) if $D_j = a_j$ is taken into account. In this way, the consistency between the stochastic differential equation (10.115) and Fokker-Planck equation (10.85) is also demonstrated regarding the implied correlation dynamics.

## 10.5 Molecular and Fluid Motion

The mathematical modeling of molecular and fluid motion is a problem that is relevant to a huge variety of processes in nature (e.g., atmospheric dynamics) and technology (e.g., reactor chemistry). Actually, we consider only one process: the motion of molecules of a fluid. The difference between the terms molecular and fluid motion is given by the scale considered. With regard to molecular motion we are interested in an understanding of elementary processes with a typical length scale of about $10^{-9}$ m, whereas the consideration of fluid dynamics means to look at processes with a typical length scale of about $10^{-3}$ m. Fluid dynamic variables represent means of molecular variables. For example, the mean molecular velocity is equal to the fluid dynamic velocity. Therefore, we will derive here the equations for fluid motion as the moment equations that are implied by a stochastic model for the molecular motion. From a mathematical point of view, the goal of this section is to illustrate the application of stochastic differential equations and the Fokker-Planck equation. In particular, the goals are to show:

- the typical structure of stochastic differential equations for a real problem,
- the problem related to the numerical solution of such stochastic equations,
- the use of analysis tools for deriving moment equations,
- the typical closure problem of moment equations,
- a consistent and systematic way to develop closed moment equations,
- ways to assess the range of validity of different moment equations.

The focus here is on the modeling problem, this means the derivation of closed equations for molecular and fluid motion. Unfortunately, the equations obtained cannot be solved analytically, and numerical solutions turn out to be extremely expensive. Interested readers may find more information about solutions of these equations elsewhere (Pope 2000, Heinz 2003, 2004, Fox 2003, Givi 2006, Jenny et al. 2010). The modeling problem will be considered here in its simplest form, this means without accounting for additional variables (like mass fractions of chemical species) or forces (like the gravity force). Such modifications, which may be relevant to applications, can be taken into account by following the methodology to be presented in the following.

### *10.5.1 Molecular Motion Model*

**Stochastic Molecular Motion Model.** Attention will be restricted here to the case of monatomic fluids, which do not have internal degrees of freedom (rotational or vibrational energy). The molecules are assumed to move independently. This corresponds to the consideration of a perfect gas. The state of each molecule is completely described by its position $x_i^*$ and velocity $V_i^*$. Here, the subscript $i = 1, 3$ indicates the three position and velocity components in physical space. The equations considered for $x_i^*$ and $V_i^*$ are given by (Heinz 2003, 2004, 2007)

$$\frac{dx_i^*}{dt} = V_i^*, \tag{10.125a}$$

$$\frac{dV_i^*}{dt} = -\frac{V_i^* - U_i}{\tau} + \sqrt{\frac{4e}{3\tau}}\frac{dW_i}{dt}. \tag{10.125b}$$

Here, $dW_i / dt$ is the derivative of a Wiener process. This model involves three parameters: $U_i$ is the mean molecular velocity, $e$ is the specific kinetic energy (it has the dimension of a squared velocity), and $\tau$ is the characteristic time scale of molecular fluctuations. All the three parameters may depend on time $t$ and the position of a molecule (the model parameters are functions of the position $\boldsymbol{x}$ in physical space, where $\boldsymbol{x}$ is replaced by $\boldsymbol{x}^*(t)$ in Eqs. (10.125)). The application of this model does only require the definition of the time scale $\tau$, because $U_i$ and $e$ can be calculated from molecular properties (by taking the mean over velocities and squared velocity fluctuations). An external force is not considered here for simplicity. The model considered represents an extension of the Brownian motion model (6.59). A difference is given by the inclusion of the mean velocity $U_i$ in the drift term here, which is assumed to be zero in the Brownian motion model (6.59). Equations (10.125) can be solved via Monte Carlo simulation, which enables the calculation of all relevant variables (like $U_i$) as means over particle properties. Nevertheless, this approach is computationally very expensive (Jenny et al. 2010). It is usually more convenient to consider equations for moments, which can be solved with lower computational cost.

**Moment Equations Implied by Stochastic Model.** Similar to the analysis of the Brownian motion model (6.59) we can study the consequences of the stochastic model (10.125) for statistical particle properties, this means we can calculate the evolution of the mean particle position, velocity, and variances in time. However, our main interest here is in fluid dynamics, i.e., the properties of the fluid at a fixed position and time. Such fluid dynamics properties are given by the fluid mass density $\rho(\boldsymbol{x}, t)$ and fluid velocity $U_i(\boldsymbol{x}, t)$. Equations for $\rho(\boldsymbol{x}, t)$ and $U_i(\boldsymbol{x}, t)$ can be derived as a consequence of the Fokker-Planck equation that is implied by

the stochastic molecular model (10.125). However, these derivations are relatively lengthy. Therefore, only the resulting equations are presented here. All the details of how these equations can be obtained are given in the appendix of this section (see Sect. 10.5.3). All the equations presented in this paragraph are exact consequences of the stochastic model (10.125). The evolution equations for $\rho(\boldsymbol{x}, t)$ and $U_i(\boldsymbol{x}, t)$ can be presented efficiently in terms of the substantial derivative $DQ/Dt = \partial Q/\partial t + U_m \partial Q/\partial x_m$ (see the discussion of this derivative in Sect. 10.1), where $Q(\boldsymbol{x}, t)$ can be any variable. The equations for $\rho(\boldsymbol{x}, t)$ and $U_i(\boldsymbol{x}, t)$ that are implied by the stochastic molecular model (10.125) can be written then

$$\frac{D\rho}{Dt} + \rho \frac{\partial U_m}{\partial x_m} = 0, \tag{10.126a}$$

$$\frac{DU_i}{Dt} + \frac{2}{3\rho}\frac{\partial \rho e}{\partial x_i} + \frac{1}{\rho}\frac{\partial \rho d_{im}}{\partial x_m} = 0. \tag{10.126b}$$

There are two unknown variables in the last equation: the kinetic energy $e(\boldsymbol{x}, t)$ and deviatoric stress $d_{ij}(\boldsymbol{x}, t)$. Both $e$ and $d_{ij}$ are related to the variance of molecular velocities. In particular, $e$ represents the isotropic variance contribution, and $d_{ij}$ is the anisotropic variance contribution (see the corresponding explanations in Sect. 10.5.3). The stochastic model (10.125) implies an equation for the variance, which can be used to derive the following equations for $e$ and $d_{ij}$,

$$\frac{De}{Dt} + \frac{1}{2\rho}\frac{\partial \rho \overline{v_i v_i v_m}}{\partial x_m} + \frac{\partial U_i}{\partial x_m} d_{mi} + \frac{2}{3} e \frac{\partial U_i}{\partial x_i} = 0, \tag{10.127a}$$

$$\begin{aligned} &\frac{Dd_{ij}}{Dt} + \frac{1}{\rho}\frac{\partial \rho \overline{(v_i v_j - v_k v_k \delta_{ij}/3) v_m}}{\partial x_m} + \frac{\partial U_i}{\partial x_m}\left(d_{mj} + \frac{2}{3} e \delta_{mj}\right) \\ &\quad + \frac{\partial U_j}{\partial x_m}\left(d_{mi} + \frac{2}{3} e \delta_{mi}\right) - \frac{2}{3}\delta_{ij}\frac{\partial U_k}{\partial x_m}\left(d_{mk} + \frac{2}{3} e \delta_{mk}\right) = -\frac{2}{\tau} d_{ij}. \end{aligned} \tag{10.127b}$$

The $d_{ij}$ equation provides zero on both sides if we set $i = j$ and take the sum over $j$. Equations (10.127) contain again unknowns given by the terms that involve three velocity fluctuations $v_i$ (the triple correlation). An equation for these triple correlations can be also derived from the stochastic model. This equation reads

$$\begin{aligned} &\frac{D\overline{v_i v_j v_k}}{Dt} + \frac{1}{\rho}\frac{\partial \rho \overline{v_i v_j v_k v_m}}{\partial x_m} - \frac{1}{\rho}\frac{\partial \rho \overline{v_i v_m}}{\partial x_m}\overline{v_j v_k} - \frac{1}{\rho}\frac{\partial \rho \overline{v_j v_m}}{\partial x_m}\overline{v_i v_k} - \frac{1}{\rho}\frac{\partial \rho \overline{v_k v_m}}{\partial x_m}\overline{v_i v_j} \\ &\quad + \frac{\partial U_i}{\partial x_m}\overline{v_m v_j v_k} + \frac{\partial U_j}{\partial x_m}\overline{v_m v_i v_k} + \frac{\partial U_k}{\partial x_m}\overline{v_m v_i v_j} = -\frac{3}{\tau}\overline{v_i v_j v_k}. \end{aligned} \tag{10.128}$$

**Discussion of Moment Equations.** It is interesting that Eqs. (10.126) for the fluid mass density $\rho(\boldsymbol{x}, t)$ and the fluid velocity $U_i(\boldsymbol{x}, t)$ are equal to equations that follow from several molecular motion equations, as, for example, the Boltzmann equation (Heinz 2003, 2004, Jenny et al. 2010). The same applies to the left-hand sides of Eqs. (10.127) and (10.128). Hence, the influence of the stochastic molecular model (10.125) considered does only appear on the right-hand sides of the $d_{ij}$ Eq. (10.127b) and triple correlation equation (10.128). Other molecular motion models provide right-hand sides of the $d_{ij}$ and triple correlation equations that have the same structure (Jenny et al. 2010). The equation system (10.126)–(10.128) of coupled fluid dynamics equations is still unclosed due to the appearance of the term with four velocity fluctuations $v_i$ in Eq. (10.128).

### *10.5.2 Fluid Dynamics Equations*

Next, let us consider how it is possible to overcome the closure problem of Eqs. (10.126)–(10.128) described in the previous paragraph, i.e., how closed equations for fluid dynamics can be derived.

**Algebraic Model for Fourth-Order Correlations.** To close Eqs. (10.126)–(10.128) we need a model for the unknown fourth-order velocity correlations in the triple correlation equation (10.128). A corresponding closure model can be obtained by assuming that the velocity PDF can be approximated by a joint normal PDF, which leads (in generalization of Eqs. (10.43b) for the fourth-order correlations of a bivariate normal distribution) to the following parametrization of fourth-order central velocity moments,

$$\overline{v_i v_j v_k v_m} = \overline{v_i v_j}\,\overline{v_k v_m} + \overline{v_i v_k}\,\overline{v_j v_m} + \overline{v_i v_m}\,\overline{v_j v_k}. \tag{10.129}$$

This approximation represents a reasonable assumption for all fluids that are not too far from an equilibrium state. It is relevant to see that this assumption does only affect the evolution of triple correlations, which are small for fluids that are close to an equilibrium state. The gradient of fourth-order correlations required in the triple correlation equation (10.128) is then given by

$$\begin{aligned}\frac{1}{\rho}\frac{\partial\rho\overline{v_i v_j v_k v_m}}{\partial x_m} = {} & \frac{\partial\overline{v_i v_j}}{\partial x_m}\overline{v_k v_m} + \frac{\partial\overline{v_i v_k}}{\partial x_m}\overline{v_j v_m} + \frac{\partial\overline{v_j v_k}}{\partial x_m}\overline{v_i v_m} \\ & + \frac{1}{\rho}\frac{\partial\rho\overline{v_k v_m}}{\partial x_m}\overline{v_i v_j} + \frac{1}{\rho}\frac{\partial\rho\overline{v_j v_m}}{\partial x_m}\overline{v_i v_k} + \frac{1}{\rho}\frac{\partial\rho\overline{v_i v_m}}{\partial x_m}\overline{v_j v_k}.\end{aligned} \tag{10.130}$$

The use of this expression in Eq. (10.128) for velocity triple correlations leads to a closed equation for triple correlations,

$$\begin{aligned}\frac{D\overline{v_i v_j v_k}}{Dt} &+ \frac{\partial \overline{v_i v_j}}{\partial x_m}\overline{v_k v_m} + \frac{\partial \overline{v_i v_k}}{\partial x_m}\overline{v_j v_m} + \frac{\partial \overline{v_j v_k}}{\partial x_m}\overline{v_i v_m} \\ &+ \frac{\partial U_i}{\partial x_m}\overline{v_m v_j v_k} + \frac{\partial U_j}{\partial x_m}\overline{v_m v_i v_k} + \frac{\partial U_k}{\partial x_m}\overline{v_m v_i v_j} = -\frac{3}{\tau}\overline{v_i v_j v_k}.\end{aligned} \tag{10.131}$$

In this way we have derived a closed system of fluid dynamics equations given by Eqs. (10.126), (10.127), and (10.131).

**Algebraic Model for Third-Order Correlations.** The cost of simulations can be reduced by using Eq. (10.131) for the derivation of an algebraic approximation for the triple correlations. We assume that the substantial derivative and the terms that contain triple correlations multiplied with velocity gradients can be neglected in comparison to the other terms,

$$\overline{v_i v_j v_k} = -\frac{\tau}{3}\left(\frac{\partial \overline{v_i v_j}}{\partial x_m}\overline{v_k v_m} + \frac{\partial \overline{v_i v_k}}{\partial x_m}\overline{v_j v_m} + \frac{\partial \overline{v_j v_k}}{\partial x_m}\overline{v_i v_m}\right). \tag{10.132}$$

This model implies for the triple correlations in the energy equation

$$\begin{aligned}\overline{v_i v_i v_m} &= -\frac{2\tau}{3}\left[\frac{\partial e}{\partial x_n}\overline{v_m v_n} + \frac{\partial \overline{v_i v_m}}{\partial x_n}\overline{v_i v_n}\right] = -\frac{2\tau}{3}\overline{v_i v_n}\,\frac{\partial\left(e\delta_{im} + \overline{v_i v_m}\right)}{\partial x_n} \\ &= -\frac{2\tau}{3}\left(\frac{2}{3}e\delta_{in} + d_{in}\right)\frac{\partial}{\partial x_n}\left(\frac{5}{3}e\delta_{im} + d_{im}\right),\end{aligned} \tag{10.133}$$

where the variances $\overline{v_i v_m} = 2\,e/3\,\delta_{im} + d_{im}$ are represented by the specific kinetic energy $e = \overline{v_i v_i}/2$ and deviatoric stress $d_{im}$ (see Sect. 10.5.3). In this case, we have a closed system of fluid dynamics equations given by Eqs. (10.126) and (10.127) combined with Eqs. (10.132) and (10.133), respectively.

**Algebraic Model for Second-Order Correlations.** The fluid dynamics equations can be further simplified by the derivation of an algebraic model for $d_{ij}$. This model can be obtained by neglecting $Dd_{ij}/Dt$, the gradients of triple correlations, and the anisotropy contributions $d_{ij}$ in the parenthesis terms of Eq. (10.127b),

$$\frac{2}{3}e\left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i} - \frac{2}{3}\delta_{ij}\frac{\partial U_k}{\partial x_k}\right) = -\frac{2}{\tau}d_{ij}. \tag{10.134}$$

This relation can be written more efficiently by introducing the shear rate tensor

$$S_{ij} = \frac{1}{2}\left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i}\right) \tag{10.135}$$

and the related deviatoric shear rate tensor

$$S_{ij}{}^{d} = S_{ij} - \frac{1}{3} S_{nn} \delta_{ij} = \frac{1}{2} \left( \frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i} - \frac{2}{3} \frac{\partial U_n}{\partial x_n} \delta_{ij} \right), \tag{10.136}$$

which is the deviation of the shear rate tensor from its isotropic part. According to its definition, $S_{ij}{}^{d}$ has the property $S_{ii}{}^{d} = 0$. In terms of the definition of $S_{ij}{}^{d}$, the algebraic model (10.134) for $d_{ij}$ can be written

$$d_{ij} = -2 \frac{\tau e}{3} S_{ij}{}^{d} = -2 \nu S_{ij}{}^{d}. \tag{10.137}$$

The last expression introduces the diffusion coefficient

$$\nu = \frac{\tau e}{3}, \tag{10.138}$$

which is called the kinematic viscosity.

**Fluid Dynamics Equations.** The use of the approximation (10.137) simplifies the fluid dynamics equations significantly. Equations (10.126) read now

$$\frac{D\rho}{Dt} = -\rho S_{ii}, \tag{10.139a}$$

$$\frac{DU_i}{Dt} = \frac{2}{\rho} \frac{\partial \rho \nu S_{im}{}^{d}}{\partial x_m} - \frac{2}{3\rho} \frac{\partial \rho e}{\partial x_i}. \tag{10.139b}$$

In the first equation we applied $S_{ii} = \partial U_i / \partial x_i$, which follows from Eq. (10.135). The second equation is often written in terms of the viscosity $\mu = \rho \nu$. The energy is given by Eq. (10.127a) combined with Eq. (10.133) for $\overline{v_i v_i v_m}$. In $\overline{v_i v_i v_m}$ we have to neglect anisotropy contributions $d_{ij}$ in the parenthesis terms to be consistent with the approximations used in the $d_{ij}$ equation – we have

$$\overline{v_i v_i v_m} = -\frac{20 \tau e}{27} \frac{\partial e}{\partial x_m} = -\frac{20}{9} \nu \frac{\partial e}{\partial x_m} = -2 \nu \frac{\gamma}{Pr} \frac{\partial e}{\partial x_m}. \tag{10.140}$$

The last writing presents this expression in its standard formulation. Here, $Pr$ is the Prandtl number, and $\gamma = 1 + 2/f$ is the ratio of specific heats, where $f$ counts the degrees of freedom. Monatomic gases have $f = 3$ degrees of freedom such that $\gamma = 5/3$. Therefore, we have $Pr = 3/2$ for the case considered (Jenny et al. 2010). By using Eq. (10.140) for the triple correlation and Eq. (10.137) for $d_{ij}$ we find the energy equation to be given by

$$\frac{De}{Dt} = \frac{1}{\rho} \frac{\partial}{\partial x_m} \left( \rho \nu \frac{\gamma}{Pr} \frac{\partial e}{\partial x_m} \right) + 2 \nu \frac{\partial U_i}{\partial x_m} S_{mi}{}^{d} - \frac{2}{3} e \frac{\partial U_i}{\partial x_i}. \tag{10.141}$$

The last two terms of this equation can be rewritten in terms of the definitions of $S_{ij}$ and $S_{ij}^{\ d}$. We use again $S_{ii} = \partial U_i / \partial x_i$. The term involving $S_{mi}^{\ d}$ can be written

$$\frac{\partial U_i}{\partial x_m} S_{mi}^{\ d} = \frac{1}{2}\left( \frac{\partial U_i}{\partial x_m} + \frac{\partial U_m}{\partial x_i} - \frac{2}{3}\frac{\partial U_k}{\partial x_k}\delta_{im} \right) S_{mi}^{\ d} = S_{mi}^{\ d}\, S_{mi}^{\ d}, \tag{10.142}$$

which is a consequence of symmetry properties and the fact that $S_{ii}^{\ d} = 0$. Thus, the energy equation can be written as

$$\frac{De}{Dt} = \frac{1}{\rho}\frac{\partial}{\partial x_m}\left( \rho \nu \frac{\gamma}{Pr}\frac{\partial e}{\partial x_m} \right) + 2\nu\, S_{mi}^{\ d}\, S_{im}^{\ d} - \frac{2}{3} e\, S_{ii}. \tag{10.143}$$

Equations (10.139) combined with this energy equation represent a closed equation system. The equations can be presented in different ways by using the relations between the kinetic energy $e$ with the pressure $p$ and temperature $T$,

$$e = \frac{3}{2}\frac{p}{\rho} = \frac{3}{2} RT. \tag{10.144}$$

Here, R refers to the gas constant.

**Navier-Stokes Equations.** Equations (10.139) combined with Eq. (10.143) represent the Navier-Stokes equations, where the ratio $\gamma / Pr$ is chosen according to the fluid considered. The value $\gamma = 5/3$ is the correct value for monatomic gases, but a Prandtl number value $Pr = 3/2$ derived here as a consequence of the simple molecular model (10.125) needs adjustments (for most gases measurements show a more or less constant Prandtl number value $Pr = 2/3$). Which influences cause changes of the fluid dynamic variables? The mass density $\rho$ is changed by the dilatation $S_{ii}$, which measures compressibility. Changes of the fluid velocity $U_i$ are caused by two effects: molecular diffusion (the first term on the right-hand side) and kinetic energy (or pressure) gradients (the last term): a decreasing pressure in the $x_i$ direction (i.e., a negative pressure gradient) implies a positive acceleration $DU_i / Dt$ of the fluid in this direction. Changes of the kinetic energy can be caused by three effects. The first effect is given by molecular diffusion (the first term on the right-hand side). The second effect is given by viscous heating (the second term). This contribution is always positive. It arises from the conversion of kinetic energy into heat. The third effect is due to compressibility (the last term). Analytical solutions of the Navier-Stokes equations can be only found under very specific conditions. Numerical solutions of the Navier-Stokes equations turn out to be extremely expensive if the fluid considered is turbulent, which is the usual case (Pope 2000). Therefore, studies of fluid properties on the basis of these equations usually represent a very complicated matter. A simple illustration of characteristic properties of the Navier-Stokes equations was given in Chap. 9 by the discussion of the Lorenz equations and their chaotic solutions.

### *10.5.3 Appendix: Implications of the Stochastic Molecular Model*

This section shows how the equations of fluid dynamics (10.126), (10.127), and (10.128) can be derived from the stochastic molecular motion model (10.125).

**Joint PDF and Conditional PDF.** The joint PDF for molecular positions $x_i^*(t)$ and velocities $V_i^*(t)$ involved in the molecular model (10.125) is defined by

$$f(\boldsymbol{w},\boldsymbol{x},t)=\left\langle \delta\left(\boldsymbol{x}^*(t)-\boldsymbol{x}\right)\delta\left(\boldsymbol{V}^*(t)-\boldsymbol{w}\right)\right\rangle, \tag{10.145}$$

where $\boldsymbol{x}$ and $\boldsymbol{w}$ refer to the sample space positions and velocities, respectively. The brackets denote an ensemble average. To define fluid dynamic variables at fixed positions $\boldsymbol{x}$ we need the conditional PDF

$$F(\boldsymbol{w},\boldsymbol{x},t)=\frac{1}{\left\langle \delta\left(\boldsymbol{x}^*(t)-\boldsymbol{x}\right)\right\rangle}\left\langle \delta\left(\boldsymbol{x}^*(t)-\boldsymbol{x}\right)\delta\left(\boldsymbol{V}^*(t)-\boldsymbol{w}\right)\right\rangle. \tag{10.146}$$

The delta function $\delta(\boldsymbol{x}^*(t)-\boldsymbol{x})$ involved here is proportional to the instantaneous molecular mass density, which is defined by

$$\rho^*(\boldsymbol{x},t)=M\,\delta\left(\boldsymbol{x}^*(t)-\boldsymbol{x}\right). \tag{10.147}$$

The integration of Eq. (10.147) over $\boldsymbol{x}$ shows that $M=\int\rho^*(\boldsymbol{x},t)\,d\boldsymbol{x}$. Hence, $M$ is the total mass of molecules within the domain considered. The mean molecular mass density is given by averaging Eq. (10.147),

$$\rho(\boldsymbol{x},t)=\left\langle \rho^*(\boldsymbol{x},t)\right\rangle. \tag{10.148}$$

By applying Eqs. (10.147) and (10.148), the conditional PDF can be written

$$F(\boldsymbol{w},\boldsymbol{x},t)=\frac{1}{\rho(\boldsymbol{x},t)}\left\langle \rho^*(\boldsymbol{x},t)\,\delta\left(\boldsymbol{V}^*(t)-\boldsymbol{w}\right)\right\rangle. \tag{10.149}$$

Hence, the joint PDF $f(\boldsymbol{w},\boldsymbol{x},t)$ and the conditional PDF $F(\boldsymbol{w},\boldsymbol{x},t)$ are related by $f(\boldsymbol{w},\boldsymbol{x},t)=\rho(\boldsymbol{x},t)\,F(\boldsymbol{w},\boldsymbol{x},t)/M$. Expression (10.149) shows that $F(\boldsymbol{w},\boldsymbol{x},t)$ integrates to one, this means $\int F(\boldsymbol{w},\boldsymbol{x},t)\,d\boldsymbol{w}=1$.

**Fluid Dynamic Variables.** Integrations over $F(\boldsymbol{w},\boldsymbol{x},t)$ provide fluid dynamic variables at a fixed position $\boldsymbol{x}$ and time $t$. For any function $Q$ of velocities we find

$$\begin{aligned}\int Q(\boldsymbol{w})\,F(\boldsymbol{w},\boldsymbol{x},t)\,d\boldsymbol{w}&=\frac{1}{\rho(\boldsymbol{x},t)}\int Q(\boldsymbol{w})\left\langle \rho^*(\boldsymbol{x},t)\,\delta\left(\boldsymbol{V}^*(t)-\boldsymbol{w}\right)\right\rangle d\boldsymbol{w}\\&=\frac{1}{\rho(\boldsymbol{x},t)}\int\left\langle \rho^*(\boldsymbol{x},t)\,Q\left(\boldsymbol{V}^*(t)\right)\delta\left(\boldsymbol{V}^*(t)-\boldsymbol{w}\right)\right\rangle d\boldsymbol{w}\\&=\frac{1}{\rho(\boldsymbol{x},t)}\left\langle \rho^*(\boldsymbol{x},t)\,Q\left(\boldsymbol{V}^*(t)\right)\right\rangle=\overline{Q(\boldsymbol{V})}(\boldsymbol{x},t).\end{aligned} \tag{10.150}$$

The first line makes use of the definition of $F(\boldsymbol{w}, \boldsymbol{x}, t)$. In the second line, $Q(\boldsymbol{w})$ is written inside the brackets and replaced by $Q(\boldsymbol{V}^*(t))$ according to the sifting property of delta functions. The normalization property of delta functions is used in the third line. The last expression introduces an abbreviation for the previous expression that refers to the physical meaning of this expression: we calculate the mass-density weighted mean over velocities at a fixed position $\boldsymbol{x}$ and time $t$ in this way. Examples for the use of Eq. (10.150) are given by the following definitions of the first three velocity moments

$$\overline{V_i}(\boldsymbol{x},t) = \int w_i \, F(\boldsymbol{w},\boldsymbol{x},t)\, d\boldsymbol{w} = U_i(\boldsymbol{x},t), \tag{10.151a}$$

$$\overline{V_i V_j}(\boldsymbol{x},t) = \int w_i w_j \, F(\boldsymbol{w},\boldsymbol{x},t)\, d\boldsymbol{w}, \tag{10.151b}$$

$$\overline{V_i V_j V_k}(\boldsymbol{x},t) = \int w_i w_j w_k \, F(\boldsymbol{w},\boldsymbol{x},t)\, d\boldsymbol{w}. \tag{10.151c}$$

Relation (10.151a) relates the integral to the mean molecular velocity $U_i$, which is used in the stochastic molecular model formulation.

**Fokker-Planck Equation.** Equations for the fluid dynamic variables (10.151) can be found as a consequence of the stochastic molecular model (10.125). The equation for the joint PDF $f(\boldsymbol{w}, \boldsymbol{x}, t)$, which is implied by the stochastic molecular model, is given by

$$\frac{\partial f}{\partial t} = -\frac{\partial w_m f}{\partial x_m} - \frac{\partial}{\partial w_m}\left[-\frac{w_m - U_m}{\tau}\right] f + \frac{2e}{3\tau}\frac{\partial^2 f}{\partial w_m \partial w_m}. \tag{10.152}$$

By multiplying this equation by $M$ and using the relation between the joint PDF and conditional PDF, $f(\boldsymbol{w}, \boldsymbol{x}, t) = \rho(\boldsymbol{x}, t)\, F(\boldsymbol{w}, \boldsymbol{x}, t) / M$, we find

$$\frac{\partial \rho F}{\partial t} + \frac{\partial \rho w_m F}{\partial x_m} = \frac{\partial}{\partial w_m}\left[\frac{w_m - U_m}{\tau}\rho F + \frac{2e}{3\tau}\frac{\partial \rho F}{\partial w_m}\right]. \tag{10.153}$$

**Mass Density Equation.** The integration of the latter equation over the velocity sample space $\boldsymbol{w}$ implies an equation for the mean mass density,

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho U_m}{\partial x_m} = \int \frac{\partial}{\partial w_m}\left[\frac{w_m - U_m}{\tau}\rho F + \frac{2e}{3\tau}\frac{\partial \rho F}{\partial w_m}\right] d\boldsymbol{w} = 0. \tag{10.154}$$

We used here $\overline{V_i}(\boldsymbol{x}, t) = U_i(\boldsymbol{x}, t)$ and the normalization property of $F$. The terms on the right-hand side do not contribute because we have integrals over derivatives, which disappear at infinity. By distributing the spatial derivative we find

$$\frac{D\rho}{Dt} + \rho \frac{\partial U_m}{\partial x_m} = 0, \tag{10.155}$$

which corresponds to Eq. (10.126a). We applied the definition of the substantial derivative $DQ/Dt = \partial Q/\partial t + U_m \partial Q/\partial x_m$ (see Sect. 10.1), where $Q(\boldsymbol{x}, t)$ can be any variable.

**Velocity Equation.** An equation for the mean velocity can be derived by multiplication of Eq. (10.153) with $w_i$ and integration over the velocity sample space,

$$\begin{aligned}\frac{\partial \rho U_i}{\partial t} + \frac{\partial \rho \overline{V_i V_m}}{\partial x_m} &= \int w_i \frac{\partial}{\partial w_m}\left[\frac{w_m - U_m}{\tau}\rho F + \frac{2e}{3\tau}\frac{\partial \rho F}{\partial w_m}\right] d\boldsymbol{w} \\ &= -\int \left[\frac{w_i - U_i}{\tau}\rho F + \frac{2e}{3\tau}\frac{\partial \rho F}{\partial w_i}\right] d\boldsymbol{w} = 0,\end{aligned} \tag{10.156}$$

where integration by parts is applied. This equation can be rewritten by splitting $\overline{V_i V_m}$ into contributions due to the mean velocity $U_i$ and deviations $v_i = V_i - U_i$ from the mean velocity,

$$\overline{V_i V_m} = U_i U_m + \overline{v_i v_m}. \tag{10.157}$$

The last term represents the variance of the velocity distribution. The consistency of this relation may be seen by distributing the variance according to $V_i = U_i + v_i$. The combination of Eq. (10.156) with Eq. (10.157) leads to the following mean velocity equation

$$\frac{\partial \rho U_i}{\partial t} + \frac{\partial \rho U_i U_m}{\partial x_m} + \frac{\partial \rho \overline{v_i v_m}}{\partial x_m} = 0. \tag{10.158}$$

For any function $Q(\boldsymbol{x}, t)$ we have the relation

$$\frac{\partial \rho Q}{\partial t} + \frac{\partial \rho Q U_m}{\partial x_m} = Q\left[\frac{\partial \rho}{\partial t} + \frac{\partial \rho U_m}{\partial x_m}\right] + \rho\left[\frac{\partial Q}{\partial t} + U_m \frac{\partial Q}{\partial x_m}\right] = \rho \frac{DQ}{Dt}. \tag{10.159}$$

The first bracket term does not contribute here because of Eq. (10.154). By setting $Q = U_i$ and using the last relation, we can write the velocity equation as

$$\frac{DU_i}{Dt} + \frac{1}{\rho}\frac{\partial \rho \overline{v_i v_m}}{\partial x_m} = 0. \tag{10.160}$$

To prepare the use of approximations (see Sect. 10.5.2) it is helpful to split the variance into two contributions, $\overline{v_i v_m} = 2e/3\ \delta_{im} + d_{im}$. Here, the kinetic energy $e$ and deviatoric stress $d_{ij}$, which has the property $d_{ii} = 0$, are given by

$$e = \frac{1}{2}\overline{v_i v_i}, \qquad d_{ij} = \overline{v_i v_j} - \frac{2}{3} e \delta_{ij}. \tag{10.161}$$

The resulting equation for $U_i$ is then equal to Eq. (10.126b),

$$\frac{DU_i}{Dt}+\frac{2}{3\rho}\frac{\partial \rho e}{\partial x_i}+\frac{1}{\rho}\frac{\partial \rho d_{im}}{\partial x_m}=0. \tag{10.162}$$

**Variance Equation.** Equation (10.160) is unclosed due to the appearance of the variance $\overline{v_i v_m}$. To derive an equation for $\overline{v_i v_m}$ we multiply Eq. (10.153) by $(w_i - U_i)$ $(w_j - U_j)$ and integrate over the velocity sample space. Let us separately calculate the right-hand side (RHS) and left-hand side (LHS) of this equation,

$$\begin{aligned}
\text{RHS} &= \int (w_i-U_i)(w_j-U_j)\frac{\partial}{\partial w_m}\left[\frac{w_m-U_m}{\tau}\rho F+\frac{2e}{3\tau}\frac{\partial \rho F}{\partial w_m}\right]d\boldsymbol{w}\\
&= -\int\left(\delta_{im}(w_j-U_j)+\delta_{jm}(w_i-U_i)\right)\left[\frac{w_m-U_m}{\tau}\rho F+\frac{2e}{3\tau}\frac{\partial \rho F}{\partial w_m}\right]d\boldsymbol{w}\\
&= -\frac{2}{\tau}\int (w_i-U_i)(w_j-U_j)\,\rho F\,d\boldsymbol{w}+\frac{2e}{3\tau}\int(\delta_{im}\delta_{jm}+\delta_{jm}\delta_{im})\,\rho F\,d\boldsymbol{w}\\
&= -\frac{2}{\tau}\rho\overline{v_i v_j}+\frac{4e}{3\tau}\rho\delta_{ij},
\end{aligned} \tag{10.163}$$

where integration by parts is applied. The term on the left-hand side is given by

$$\begin{aligned}
\text{LHS} &= \int (w_i-U_i)(w_j-U_j)\left[\frac{\partial \rho F}{\partial t}+\frac{\partial \rho U_m F}{\partial x_m}+\frac{\partial \rho (w_m-U_m)F}{\partial x_m}\right]d\boldsymbol{w}\\
&= \frac{\partial \rho \overline{v_i v_j}}{\partial t}+\frac{\partial \rho U_m \overline{v_i v_j}}{\partial x_m}+\frac{\partial \rho \overline{v_i v_j v_m}}{\partial x_m}+\rho\frac{\partial U_i}{\partial x_m}\overline{v_m v_j}+\rho\frac{\partial U_j}{\partial x_m}\overline{v_m v_i}\\
&= \rho\frac{D\overline{v_i v_j}}{Dt}+\frac{\partial \rho \overline{v_i v_j v_m}}{\partial x_m}+\rho\frac{\partial U_i}{\partial x_m}\overline{v_m v_j}+\rho\frac{\partial U_j}{\partial x_m}\overline{v_m v_i}.
\end{aligned} \tag{10.164}$$

This rewriting is obtained by using the derivatives first such that they apply to all the integral and adding then corrections (given by the terms that involve velocity gradients). The last expression results from the use of Eq. (10.159). The combination of Eqs. (10.163) and (10.164) leads then to the variance equation

$$\frac{D\overline{v_i v_j}}{Dt}+\frac{1}{\rho}\frac{\partial \rho \overline{v_i v_j v_m}}{\partial x_m}+\frac{\partial U_i}{\partial x_m}\overline{v_m v_j}+\frac{\partial U_j}{\partial x_m}\overline{v_m v_i}=-\frac{2}{\tau}\left(\overline{v_i v_j}-\frac{2e}{3}\delta_{ij}\right). \tag{10.165}$$

To prepare the use of approximations we split the variance into an isotropic and deviatoric part, $\overline{v_i v_j} = 2e/3\ \delta_{ij} + d_{ij}$. For $e$, Eq. (10.165) implies the equation

$$\frac{De}{Dt}+\frac{1}{2\rho}\frac{\partial \rho \overline{v_i v_i v_m}}{\partial x_m}+\frac{\partial U_i}{\partial x_m}\left(d_{mi}+\frac{2}{3}e\,\delta_{mi}\right)=0. \tag{10.166}$$

This equation corresponds to the energy equation (10.127a). Equation (10.127b) for $d_{ij}$ can be obtained by differentiating the $d_{ij}$ definition (10.161) and replacing the total derivatives of the variance and $e$ according to Eqs. (10.165) and (10.166),

$$\begin{aligned}\frac{Dd_{ij}}{Dt} &= -\frac{1}{\rho}\frac{\partial\rho\overline{v_iv_jv_m}}{\partial x_m} - \frac{\partial U_i}{\partial x_m}\left(d_{mj}+\frac{2}{3}e\,\delta_{mj}\right) - \frac{\partial U_j}{\partial x_m}\left(d_{mi}+\frac{2}{3}e\,\delta_{mi}\right) - \frac{2}{\tau}d_{ij} \\ &\quad + \frac{2}{3}\delta_{ij}\left[\frac{1}{2\rho}\frac{\partial\rho\overline{v_kv_kv_m}}{\partial x_m} + \frac{\partial U_k}{\partial x_m}\left(d_{mk}+\frac{2}{3}e\,\delta_{mk}\right)\right] \\ &= -\frac{1}{\rho}\frac{\partial\rho\overline{(v_iv_j - v_kv_k\delta_{ij}/3)v_m}}{\partial x_m} - \frac{\partial U_i}{\partial x_m}\left(d_{mj}+\frac{2}{3}e\,\delta_{mj}\right) \\ &\quad - \frac{\partial U_j}{\partial x_m}\left(d_{mi}+\frac{2}{3}e\,\delta_{mi}\right) + \frac{2}{3}\delta_{ij}\frac{\partial U_k}{\partial x_m}\left(d_{mk}+\frac{2}{3}e\,\delta_{mk}\right) - \frac{2}{\tau}d_{ij}.\end{aligned} \tag{10.167}$$

**Triple Correlation Equation.** The last two equations are unclosed due to the term that involves three velocity components. To derive an equation for this triple correlation we multiply Eq. (10.153) by $(w_i - U_i)$ $(w_j - U_j)$ $(w_k - U_k)$ and integrate this equation over the sample space. The right-hand side of this equation reads

$$\begin{aligned}\text{RHS} &= \int (w_i-U_i)(w_j-U_j)(w_k-U_k)\frac{\partial}{\partial w_m}\left[\frac{w_m-U_m}{\tau}\rho F + \frac{2e}{3\tau}\frac{\partial\rho F}{\partial w_m}\right]d\boldsymbol{w} \\ &= -\int\Big(\delta_{im}(w_j-U_j)(w_k-U_k) + \delta_{jm}(w_i-U_i)(w_k-U_k) \\ &\quad + \delta_{km}(w_i-U_i)(w_j-U_j)\Big)\left[\frac{w_m-U_m}{\tau}\rho F + \frac{2e}{3\tau}\frac{\partial\rho F}{\partial w_m}\right]d\boldsymbol{w} \\ &= -\frac{3}{\tau}\int (w_i-U_i)(w_j-U_j)(w_k-U_k)\,\rho F\,d\boldsymbol{w} = -\frac{3}{\tau}\rho\overline{v_iv_jv_k},\end{aligned} \tag{10.168}$$

where integration by parts is used. The corresponding left-hand side of the equation considered is given by

$$\begin{aligned}\text{LHS} &= \int (w_i-U_i)(w_j-U_j)(w_k-U_k)\left[\frac{\partial\rho F}{\partial t} + \frac{\partial\rho U_m F}{\partial x_m} + \frac{\partial\rho(w_m-U_m)F}{\partial x_m}\right]d\boldsymbol{w} \\ &= \frac{\partial\rho\overline{v_iv_jv_k}}{\partial t} + \frac{\partial\rho U_m\overline{v_iv_jv_k}}{\partial x_m} + \frac{\partial\rho\overline{v_iv_jv_kv_m}}{\partial x_m} + \rho\frac{DU_i}{Dt}\overline{v_jv_k} + \rho\frac{DU_j}{Dt}\overline{v_iv_k} \\ &\quad + \rho\frac{DU_k}{Dt}\overline{v_iv_j} + \rho\frac{\partial U_i}{\partial x_m}\overline{v_mv_jv_k} + \rho\frac{\partial U_j}{\partial x_m}\overline{v_mv_iv_k} + \rho\frac{\partial U_k}{\partial x_m}\overline{v_mv_iv_j}.\end{aligned} \tag{10.169}$$

The first three terms appear as a consequence of applying the derivatives to all the integral, and the remaining terms are corrections. By using Eq. (10.159) for the first two terms and Eq. (10.160) for the substantial derivatives of velocities we find the expression

$$\begin{aligned}\mathrm{LHS} = \rho\frac{D\overline{v_i v_j v_k}}{Dt} + \frac{\partial\rho\overline{v_i v_j v_k v_m}}{\partial x_m} - \frac{\partial\rho\overline{v_i v_m}}{\partial x_m}\overline{v_j v_k} - \frac{\partial\rho\overline{v_j v_m}}{\partial x_m}\overline{v_i v_k} \\ - \frac{\partial\rho\overline{v_k v_m}}{\partial x_m}\overline{v_i v_j} + \rho\frac{\partial U_i}{\partial x_m}\overline{v_m v_j v_k} + \rho\frac{\partial U_j}{\partial x_m}\overline{v_m v_i v_k} + \rho\frac{\partial U_k}{\partial x_m}\overline{v_m v_i v_j}.\end{aligned} \tag{10.170}$$

The combination of this expression with the RHS (10.168) then implies

$$\begin{aligned}\frac{D\overline{v_i v_j v_k}}{Dt} + \frac{1}{\rho}\frac{\partial\rho\,\overline{v_i v_j v_k v_m}}{\partial x_m} - \frac{1}{\rho}\frac{\partial\rho\overline{v_i v_m}}{\partial x_m}\overline{v_j v_k} - \frac{1}{\rho}\frac{\partial\rho\overline{v_j v_m}}{\partial x_m}\overline{v_i v_k} - \frac{1}{\rho}\frac{\partial\rho\overline{v_k v_m}}{\partial x_m}\overline{v_i v_j} \\ + \frac{\partial U_i}{\partial x_m}\overline{v_m v_j v_k} + \frac{\partial U_j}{\partial x_m}\overline{v_m v_i v_k} + \frac{\partial U_k}{\partial x_m}\overline{v_m v_i v_j} = -\frac{3}{\tau}\overline{v_i v_j v_k},\end{aligned} \tag{10.171}$$

which agrees with the triple correlation equation (10.128).

## 10.6 Summary

The methodological basis for the modeling of distributions of random variables and the evolution of PDFs and stochastic processes was presented for one random variable in Chaps. 4, 6, and 8. In this chapter, we extended these concepts to the case of joint random variables. Let us summarize the features observed regarding the extension of data analysis concepts, PDF modeling concepts, and concepts for describing the evolution of PDFs and stochastic processes.

**Extension of Data Analysis Concepts.** The characterization of the properties of several random variables differs from the analysis of single variable properties by the need to account for correlations (uncorrelated variables can be treated like single variables). An efficient way to account for such correlations is the use of conditional PDFs, which are rescaled joint PDFs, and related conditional means. The advantage of these concepts was demonstrated regarding the optimization of models considered in Chap. 2: a conditional mean was shown to represent an optimal model, $y_M(x) = \langle Y \mid x \rangle$. In addition to the approach presented in Chap. 2, this relation enables the development of optimal models by the calculation of the conditional mean on the basis of data, this means without the use of any modeling concepts. Other illustrations of the benefits of conditional moments can be found, e.g., in Klimenko & Bilger (1999) with regard to turbulent combustion problems.

**Extension of PDF Modeling Concepts.** Which modeling concepts can be used to describe the joint statistics of correlated variables? This is a nontrivial question, because many PDF types for single variables cannot be extended straightforwardly to the case of several variables. Here, the most relevant case was considered: it was shown that the normal PDF model for a single variable can be extended to a joint normal PDF model for several random variables that accounts correctly for any correlations. The applicability of this concept to any case considered can be proven in two ways: by showing that the normal PDF moment relations (10.43) are satisfied, or by demonstrating that scatter plots of the joint PDF agree with the consequence of a joint normal PDF (given by the elliptical shape of isolines in the ($x'$, $y'$)-coordinate system). The first way is helpful for showing that models (like the Brownian motion model considered in Sect. 6.4) have a joint normal PDF. The second way is usually applied for analyzing the joint PDF of real data (like the atmospheric velocity and temperature statistics discussed in Sect. 4.5). The joint normal PDF model does often provide the basis for modeling concepts. This was illustrated here by means of two examples: First, it was shown that the joint normal PDF model justifies the use of linear optimal models. Second, the formulation of a random walk as a sum of jointly normally distributed contributions was shown to represent a sound model: it implies a random walk process that evolves normally distributed in time, which is the typical feature of a diffusion process.

**Extension of PDF Evolution Concepts.** How is it possible to extend concepts for the evolution of PDFs of single variables to the case of several variables? It was shown that the Fokker-Planck equation for the PDF evolution and stochastic differential equation discussed in Chap. 8 can be extended to the case of several variables. As given for the single-variable case there exists a unique relationship between the Fokker-Planck equation and stochastic differential equation provided the coefficient of the noise term in the stochastic equation is a symmetric matrix. This relationship is helpful for the numerical Monte Carlo solution of diffusion-type partial differential equations that cannot be properly solved on the basis of other solution techniques. Consistent with the corresponding finding for single variables, it was shown that the Fokker-Planck equation for several variables can be solved analytically if linear dynamics of random variables are considered. The application of the PDF evolution equation presented here to the modeling of fluid dynamics in Sect. 10.5 illustrated the typical structure of stochastic models for a real problem and the typical problems related to the calculation of the solution of such equations: the numerical solution of the PDF evolution equation via Monte Carlo simulation is computationally expensive, and moment evolution equations are unclosed due to the appearance of higher-order correlations. A consistent and systematic solution for such closure problems was demonstrated by the derivation of closure models that are based on the PDF evolution equation.

## 10.7 Exercises

**10.2.1** Use the definition $f(x, y) = \langle \delta(x - X)\, \delta(y - Y) \rangle$ of a joint PDF to show that every joint PDF $f(x, y)$ of unbounded variables $x$ and $y$ has the following properties. Here, $g(x, y)$ can be any function of $x$ and $y$.

$a)\ f(x,y) \geq 0,$

$b)\ f(-\infty, y) = f(\infty, y) = f(x, -\infty) = f(x, \infty) = 0,$

$c)\ \iint f(x,y)\, dx\, dy = 1,$

$d)\ \iint g(x,y)\ f(x,y)\, dx\, dy = \big\langle g(X,Y) \big\rangle .$

**10.2.2** Consider the definition of the conditional mean

$$\big\langle g(X,Y) \,|\, x \big\rangle = \frac{1}{f(x)} \big\langle g(X,Y)\, \delta(x - X) \big\rangle .$$

Specify this definition for the case that $X$ and $Y$ are independent variables.

**10.2.3** Consider the optimal model $y_M(x) = \langle Y \,|\, x \rangle$, which was derived in Sect. 10.2.3. Assume that the joint PDF $f(x, y)$ of any data set is available as the result of measurements.

a) Explain how the optimal model $y_M(x)$ can be calculated on this basis.

b) Explain the difference between this approach for developing an optimal model and the approach applied in Chap. 2 to find an optimal model.

**10.2.4** A stochastic model for $Y$, which provides the correct mean $\langle Y \rangle$ and conditional mean $\langle Y | x \rangle$, is given by

$$Y = \big\langle Y \big\rangle + r_{XY} \big\langle \widetilde{Y}^2 \big\rangle^{1/2} \frac{X - \big\langle X \big\rangle}{\big\langle \widetilde{X}^2 \big\rangle^{1/2}}.$$

a) Calculate $\langle \widetilde{Y}^2 \rangle$ and $\langle \widetilde{X}\, \widetilde{Y} \rangle$ on the basis of this stochastic model for $Y$.

b) Use the results for $\langle \widetilde{Y}^2 \rangle$ and $\langle \widetilde{X}\, \widetilde{Y} \rangle$ to explain under which condition the model for $Y$ can represent a reasonable model.

**10.3.1** Consider the model (10.39) for the joint PDF $f(x, y)$.

a) Integrate $f(x, y)$ to show that $\int f(x, y)\, dy = f(x)$.

b) Use this result to explain why the condition $\int f(x, y)\, dx = f(y)$ is satisfied.

**10.3.2** The conditional PDF $f(y \,|\, x)$ is given by Eq. (10.47). Considered as a function of $\hat{y}$, $f(y \,|\, x)$ is a normal PDF with mean $r_{XY}\ \hat{x}$ and variance $1 - r_{XY}^2$, which is divided by $\langle \widetilde{Y}^2 \rangle^{1/2}$. Use this fact and the known properties of a normal PDF to show that $\int f(y \,|\, x)\, dy = 1$.

**10.3.3** Consider the conditional PDF $f(y\,|\,x)$ given by Eq. (10.53).

a) According to Eq. (10.19), the global variance is defined by multiplying the conditional variance $<[Y-<Y\,|\,x>]^2\,|\,x>$ with the PDF $f(x)$ and integrating over the sample space $x$. Show that this global variance is equal to the global variance $<[Y-<Y\,|\,x>_{x=X}]^2>$ considered in Sect. 10.2.3 (see the error (10.38)).

b) Calculate the conditional variance $<[Y-<Y\,|\,x>]^2\,|\,x>$ and the global variance $<[Y-<Y\,|\,x>_{x=X}]^2>$ as functions of $r_{XY}$.

c) Explain why the conditional variance $<[Y-<Y\,|\,x>]^2\,|\,x>$ is found to be equal to the global variance $<[Y-<Y\,|\,x>_{x=X}]^2>$.

**10.3.4** Consider the ellipse equation $x'^2/a^2 + y'^2/b^2 = 1$. Here, $a$ and $b$ are given by Eq. (10.60). Specify the ellipse equation for $r_{XY} \to 1$ and $r_{XY} \to -1$.

**10.3.5** The table shows the correlation coefficients of velocity components ($u$ and $v$ are horizontal velocities and $w$ is the vertical velocity) and the temperature $T$. The data were obtained by measurements for different stabilities in the atmospheric surface layer (see the discussion in Sect. 4.5).

| | $r_{uv}$ | $r_{uw}$ | $r_{vw}$ | $r_{uT}$ | $r_{vT}$ | $r_{wT}$ |
|---|---|---|---|---|---|---|
| **Stable Case:** | –0.66 | –0.04 | –0.11 | 0.84 | –0.71 | 0.02 |
| **Neutral Case:** | –0.01 | –0.26 | 0.18 | 0.39 | 0.18 | –0.18 |
| **Unstable Case:** | 0.16 | –0.13 | –0.02 | 0.01 | –0.11 | 0.50 |

a) Identify one case that is basically characterized by horizontal motions. Explain your reasoning.

b) Identify one case that indicates significant upward motions of warm air. Explain your reasoning.

c) Explain for each of the three cases considered which variables have to be accounted for in a stochastic model that characterizes the most basic features of the flow considered.

**10.3.6** According to Eq. (10.65), the PDF $f(z)$ of the sum $Z = X + Y$ of any two random variables $X$ and $Y$ is given by $f(z) = \int f(x, z-x)\,dx$.

a) The joint PDF $f(x, y)$ is assumed to be the normal PDF of two correlated variables $X$ and $Y$. Show that $f(z)$ is given for this case by

$$f(z) = \frac{\exp\left\{-\dfrac{\left(z-\langle X\rangle-\langle Y\rangle\right)^2}{2\left(\langle\widetilde{Y}^2\rangle+2\langle\widetilde{X}\,\widetilde{Y}\rangle+\langle\widetilde{X}^2\rangle\right)}\right\}}{\sqrt{2\pi\left(\langle\widetilde{Y}^2\rangle+2\langle\widetilde{X}\,\widetilde{Y}\rangle+\langle\widetilde{X}^2\rangle\right)}}.$$

b) Replace in the $f(z)$ formula the statistics of $X$ and $Y$ by the statistics of $Z$. Explain the meaning of this rewriting.

c) A nonzero correlation $< \widetilde{X}\,\widetilde{Y} >$ may lead to a lower variance of $Z$ than given for the case of uncorrelated variables $X$ and $Y$ (for which we have $< \widetilde{X}\,\widetilde{Y} > = 0$). How is it possible to understand this observation?

**10.3.7** Consider two independent random variables $X$ and $Y$, which are uniformly distributed on the interval [0, 1].

a) Calculate the PDF of the sum $Z = X + Y$.

b) Compare the result obtained in a) with the conclusions of Sect. 4.4.3 (see Fig. 4.14). What will be the PDF of a sum of a large number of independent variables that are uniformly distributed on the interval [0, 1]?

**10.4.1** The principal invariants of the symmetric matrix $D_{ij}$ are the following once:

$$I = D_{ii},$$

$$II = \frac{1}{2}\left[D_{ii}D_{nn} - D_{in}D_{ni}\right],$$

$$III = \frac{1}{6}D_{ii}D_{nn}D_{kk} - \frac{1}{2}D_{ii}D_{kn}D_{nk} + \frac{1}{3}D_{in}D_{nk}D_{ki} = \det(D).$$

a) Calculate the three principal invariants in principal axes as functions of the eigenvalues $\lambda_1$, $\lambda_2$, and $\lambda_3$. Show that $I$, $II$, and $III$ are positive if the eigenvalues $\lambda_1$, $\lambda_2$, and $\lambda_3$ are positive.

b) The three principal invariants and eigenvalues are related via the cubic characteristic equation $\lambda^3 - I\,\lambda^2 + II\,\lambda - III = 0$. Use this equation to show that the eigenvalues are positive if $I$, $II$, and $III$ are positive.

**10.4.2** Show for any matrix $\beta_{ij}(t)$ the validity of the relation

$$\frac{d\beta^{-1}{}_{ij}}{dt} = -\beta^{-1}{}_{ik}\frac{d\beta_{kn}}{dt}\beta^{-1}{}_{nj},$$

which will be applied in exercise 10.4.4. Hint: differentiate $\beta_{kn}\,\beta^{-1}{}_{nj} = \delta_{kj}$.

**10.4.3** Show for any symmetric matrix $\beta_{ij}(t)$ the validity of the relation

$$\frac{1}{\det(\beta)}\frac{d\det(\beta)}{dt} = \beta^{-1}{}_{ik}\frac{d\beta_{ki}}{dt},$$

which will be used in exercise 10.4.4. The validity of the latter relation can be shown by considering $\partial f(\boldsymbol{x}, t)/\partial t$ of the joint normal PDF

$$f(\boldsymbol{x},t) = \frac{1}{(2\pi)^{N/2}\sqrt{\det(\beta)}}\exp\left\{-\frac{1}{2}\beta^{-1}{}_{ij}\left(x_i - \alpha_i\right)\left(x_j - \alpha_j\right)\right\}$$

and integrating $\partial f(\boldsymbol{x}, t)/\partial t$ over the sample space $\boldsymbol{x}$. Here, $\alpha_i(t)$ are the mean values and $\beta_{ij}(t)$ represent the elements of the symmetric variance matrix. Hint: you have to use the relation shown in exercise 10.4.2.

**10.4.4** Consider the Fokker-Planck equation (10.107) for the conditional PDF $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$ combined with the initial condition $f(\boldsymbol{x}, t' \mid \boldsymbol{x}', t') = \delta(\boldsymbol{x} - \boldsymbol{x}')$.

a) Calculate the partial derivatives of the conditional PDF (10.109), which appear in Eq. (10.107). Hint: use the relations shown in exercises 10.4.2 and 10.4.3 to simplify the analyses in b) and c).

b) Show the conditions under which the conditional PDF (10.109) satisfies the Fokker-Planck equation (10.107).

c) Show the conditions under which the conditional PDF (10.109) satisfies the initial condition $f(\boldsymbol{x}, t' \mid \boldsymbol{x}', t') = \delta(\boldsymbol{x} - \boldsymbol{x}')$.

**10.4.5** Consider the relationship between the Fokker-Planck equation (10.85) and stochastic differential equation (10.115) discussed in Sect. 10.4.4.

a) Use this relationship to determine the evolution equation for means that is implied by the stochastic differential equation (10.115).

b) Use this relationship to find the evolution equation for variances that is implied by the stochastic equation (10.115). Write the model parameters in this equation in dependence on the stochastic process $\boldsymbol{X}(t)$ and $t$.

**10.4.6** Consider the relationship between the Fokker-Planck equation (10.85) and stochastic differential equation (10.115) discussed in Sect. 10.4.4.

a) Explain for which purpose it is particularly helpful to use the stochastic differential equation (10.115).

b) Explain for which purpose it is particularly helpful to apply the Fokker-Planck equation (10.85).

**10.5.1** Consider the stochastic velocity model (10.125). We assume that $U_i$, $e$, and $\tau$ are constants.

a) Find the equation for the velocity variance $< \widetilde{V}_i^*(t) \widetilde{V}_k^*(t) >$.

b) Solve this variance equation.

c) Explain the characteristic features of velocity variances as $t \to \infty$.

**10.5.2** Consider the stochastic velocity model (10.125). We assume that $U_i$, $e$, and $\tau$ are constants.

a) Find the equation for velocity correlations $< \widetilde{V}_i^*(t) \widetilde{V}_k^*(t+r) >$, where $r$ is any non-negative time.

b) Solve the equation for velocity correlations $< \widetilde{V}_i^*(t) \widetilde{V}_k^*(t+r) >$.

c) Calculate the velocity correlations of unequal velocity components (this means for $i \neq k$) for $t \to \infty$ by taking reference to the results obtained in exercise 10.5.1.

**10.5.3** The stochastic model (10.125) for molecular velocities provides uncoupled equations for the components of velocity. On the other hand, the variance equations (10.165), which are implied by the stochastic molecular velocity model (10.125), predict couplings between all velocity components (i.e., nonzero cross variances). Let us assume that the velocity variances were isotropic at any initial time: explain why the variances may become anisotropic after some time (what is the reason for the development of couplings between different velocity components?).

**10.5.4** The table shows fourth-order moments of wind velocity components ($u$ and $v$ are horizontal velocities, and $w$ is the vertical velocity) measured in the atmospheric surface layer for a neutral stratification. 50,400 sample values are available: see the description of these measurements in Sect. 4.5. The corresponding value found by using the normal parametrization (10.129) of fourth-order moments is given in the $\langle\rangle_N$ column. The ratio of fourth-order moments to the corresponding normal parametrization value (10.129) is shown in the $\langle\rangle / \langle\rangle_N$ column.

| | | $\langle\rangle_N$ | $\langle\rangle/\langle\rangle_N$ | | | $\langle\rangle_N$ | $\langle\rangle/\langle\rangle_N$ |
|---|---|---|---|---|---|---|---|
| $\langle u^4 \rangle$ | = 0.9241 | 0.8902 | 1.0384 | $\langle u\, v\, w^2 \rangle$ | = –0.0045 | –0.0053 | 0.8497 |
| $\langle u^3\, v \rangle$ | = –0.0299 | –0.0093 | 3.2212 | $\langle u\, w^3 \rangle$ | = –0.0323 | –0.0260 | 1.2408 |
| $\langle u^3\, w \rangle$ | = –0.1175 | –0.1116 | 1.0529 | $\langle v^4 \rangle$ | = 0.2612 | 0.2295 | 1.1381 |
| $\langle u^2\, v^2 \rangle$ | = 0.1625 | 0.1507 | 1.0782 | $\langle v^3\, w \rangle$ | = 0.0270 | 0.0281 | 0.9624 |
| $\langle u^2\, v\, w \rangle$ | = 0.0177 | 0.0192 | 0.9196 | $\langle v^2\, w^2 \rangle$ | = 0.0428 | 0.0374 | 1.1433 |
| $\langle u^2\, w^2 \rangle$ | = 0.0886 | 0.0785 | 1.1286 | $\langle v\, w^3 \rangle$ | = 0.0129 | 0.0129 | 0.9981 |
| $\langle u\, v^3 \rangle$ | = –0.0051 | –0.0047 | 1.0917 | $\langle w^4 \rangle$ | = 0.0608 | 0.0484 | 1.2562 |
| $\langle u\, v^2\, w \rangle$ | = –0.0200 | –0.0193 | 1.0364 | | | | |

a) Comment on the accuracy of data by taking reference to the number of available samples.
b) Calculate the mean value of all $\langle\rangle/\langle\rangle_N$ values by neglecting the $\langle u^3\, v \rangle$ value.
c) Do these data provide support for the normal parametrization (10.129), which was used for the derivation of fluid dynamics equations?

**10.5.5** Consider the stochastic velocity model (10.125). Assume that the positions $x_i^*$ and velocities $V_i^*$ are combined to a six-dimensional vector $\boldsymbol{Z} = (\boldsymbol{x}^*, \boldsymbol{V}^*)$. In matrix notation, Eqs. (10.125) can be written then

$$\frac{d\boldsymbol{Z}}{dt} = \boldsymbol{a} + \boldsymbol{G}(\boldsymbol{Z} - \langle \boldsymbol{Z} \rangle) + \boldsymbol{b}\frac{d\boldsymbol{W}}{dt}.$$

Here, $\boldsymbol{a}$ is a six-dimensional vector, and $\boldsymbol{G}$ and $\boldsymbol{b}$ are $6 \times 6$ matrices.
a) Specify $\boldsymbol{a}$, $\boldsymbol{G}$, and $\boldsymbol{b}$ according to the equation system (10.125).

b) What are the requirements for the coefficients $\boldsymbol{a}$, $\boldsymbol{G}$, and $\boldsymbol{b}$ under which the solution approach for Fokker-Planck equations, which was described in Sect. 10.4.3, can be used for the calculation of the velocity-position joint PDF $f(\boldsymbol{w}, \boldsymbol{x}, t)$ related to Eq. (10.125)?

**10.5.6** Consider the stochastic molecular velocity model (10.125). The asymptotic change $dV_i^* / dt$ can be considered to be small compared to the right-hand side of Eq. (10.125b). The asymptotic velocities $V_i^* = dx_i^* / dt$ are described for this case by the equation

$$\frac{dx_i^*}{dt} = U_i + \sqrt{\frac{4}{3} e \tau} \frac{dW_i}{dt}.$$

The position PDF $f(\boldsymbol{x}, t)$, which is related to $\boldsymbol{x}^*(t)$, and the conditional PDF $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$ are related by $f(\boldsymbol{x}, t) = \int f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')\, \mathrm{f}(\boldsymbol{x}', t')\, d\boldsymbol{x}'$.

a) Determine $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$ by applying the solution approach for a Fokker-Planck equation described in Sect. 10.4.3. It is assumed that $\boldsymbol{U}$, $e$, and $\tau$ are constants. Simplify $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$ as much as possible by using the expressions obtained for the parameters of $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$.

b) Determine the asymptotic conditional PDF $f(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$ as $t \to \infty$.

c) Calculate the corresponding asymptotic position PDF $f(\boldsymbol{x}, t)$.

# References


Abbot, I. H. & Von Doenhoff, A. E. 1959 Theory of Wing Sections: Including a Summary of Airfoil Data. Dover Publ., New York.

Abramowitz, M. & Stegun, I. A. 1984 Pocketbook of Mathematical Functions. Verlag Harri Deutsch, Thun, Frankfurt/Main.

Allen, L. J. S. 2003 An Introduction to Stochastic Processes with Application to Biology. Pearson Prentice Hall, Upper Saddle River, NJ.

Allen, L. J. S. 2007 An Introduction to Mathematical Biology. Pearson Prentice Hall, Upper Saddle River, NJ.

Anderson, R. M. (Editor) 1982 The Population Dynamics of Infectious Diseases: Theory and Applications. Chapman and Hall, London, New York.

Anderson, R. M. & May, R. M. 1979a Population Biology of Infectious Diseases. Part I. Nature 280, 361–367.

Anderson, R. M. & May, R. M. 1979b Population Biology of Infectious Diseases. Part 1I. Nature 280, 455–461.

Baines, P. G. 2008 Lorenz, EN 1963: Deterministic Nonperiodic Flow. J. Atmos. Sci. 20, 130–141. Prog. Phys. Geog. 32, 475–480.

Bakker, V. J., Doak, D. F., Roemer, G. W., Garcelon, D. K., Coonan, T. J., Morrison, S. A., Lynch, C., Ralls, K. & Shaw, R. 2009 Incorporating Ecological Drivers and Uncertainty into a Demographic Population Viability Analysis for the Island Fox. Ecol. Monogr. 79, 77–108.

Boyce, W. E. & DiPrima, R. C. 2009 Elementary Differential Equations and Boundary Value Problems. Ninth Edition, Wiley, Hoboken, NY.

Brannan, J. R. & Boyce, W. E. 2007 Differential Equations: An Introduction to Modern Methods and Applications. Wiley, Hoboken, NY.

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B. & Jones, P. D. 2006 Uncertainty Estimates in Regional and Global Observed Temperature Changes: A New Dataset from 1850. J. Geophys. Res. 111, D12106/1–21.


S. Heinz, *Mathematical Modeling*, DOI 10.1007/978-3-642-20311-4,

Buchanan, J. R. 2008 An Undergraduate Introduction to Financial Mathematics. Second Edition, World Scientific Publ., Singapore.

Buckingham, E. 1914 On Physically Similar Systems; Illustrations of the Use of Dimensional Equations. Phys. Rev. 4, 345–376.

Bulmer, M. G. 1979 Principles of Statistics. Second Edition, Dover Publ., New York.

Chenlo, F., Moreira, R., Pereira, G. & Bello, B. 2004 Kinematic Viscosity and Water Activity of Aqueous Solutions of Glycerol and Sodium Chloride. Eur. Food Res. Technol. 219, 403–408.

Chu, C. R., Parlange, M. B., Katul, G. G. & Albertson, J. B. 1996 Probability Density Functions of Turbulent Velocity and Temperature in the Atmospheric Surface Layer. Water Resour. Res. 32, 1681–1688.

Darwin, C. R. 1859 On the Origin of Species by Means of Natural Election. Murray, London.

Del Grosso, V. A. & Mader, C. W. 1972 Speed of Sound in Pure Water. J. Acoust. Soc. Am. 52, 1442–1446.

Du, S., Wilson, J. D. & Yee, E. 1994 Probability Density Functions for Velocity in the Convective Boundary Layer, and Implied Trajectory Models. Atmos. Environ. 28, 1211–1217.

Durbin, P. A. 1983 Stochastic Differential Equations and Turbulent Dispersion. NASA Reference Publ. 1103, NASA Glenn Research Center, Ohio.

Edelstein-Keshet, L. 2005 Mathematical Models in Biology. SIAM, Philadelphia.

Einstein, A. 1905 Über die von der Molekular-Kinetischen Theorie der Wärme Geforderte Bewegung von in Ruhenden Flüssigkeiten Suspendierten Teilchen. Ann. Phys. (Leipzig) 17, 549–560.

Feigenbaum, M. J. 1978 Quantitative Universality for a Class of Nonlinear Oscillations. J. Stat. Phys. 19, 25–52.

Fokker, A. D. 1914 Die Mittlere Energie Rotierender Elektrischer Dipole im Strahlungsfeld. Ann. Physik 43, 810–820.

Fox, R. O. 2003 Computational Models for Turbulent Reacting Flows. Cambridge Series in Chemical Engineering, Cambridge Univ. Press, Cambridge.

Fulford, G., Forrester, P. & Jones, A. 1997 Modelling with Differential and Difference Equations. Cambridge Univ. Press, Cambridge.

Gardiner, C. W. 1983 Handbook of Stochastic Methods. Springer-Verlag, Berlin, Heidelberg, New York.

Gillies, G. T. 1997 The Newtonian Gravitational Constant: Recent Measurements and Related Studies. Rep. Prog. Phys. 60, 151–225.

Giordano, F. R., Weir, M. D. & Fox, W. P. 2003 Mathematical Modeling. Third Edition, Brooks/Cole–Thomson, Pacific Grove, CA.

Givi, P. 2006 Filtered Density Function for Subgrid Scale Modeling of Turbulent Combustion. AIAA Journal 44, 16–23.

Gleick, J. 1987 Chaos: Making a New Science. Viking, New York.

Haberman, R. 1977 Mathematical Models: Mechanicals Vibrations, Population Dynamics, and Traffic Flow. Prentice-Hall, Englewood Cliffs, NJ.

Haworth, D. C. 2010 Progress in Probability Density Function Methods for Turbulent Reacting Flows. Prog. Energ. Combust. 36, 168–259.

Heinz, S. 2003 Statistical Mechanics of Turbulent Flows. Springer-Verlag, Berlin, Heidelberg, New York.

Heinz, S. 2004 Molecular to Fluid Dynamics: The Consequences of Stochastic Molecular Motion. Phys. Rev. E 70, 036308/1–11.

Heinz, S. 2007 Unified Turbulence Models for LES and RANS, FDF and PDF Simulations. Theor. Comput. Fluid Dyn. 21, 99–118.

Jaynes, E. T. 1957 Information Theory and Statistical Mechanics. Phys. Rev. 106, 620–630.

Jenny, P., Torrilhon, M. & Heinz, S. 2010 A Solution Algorithm for the Fluid Dynamic Equations Based on a Stochastic Model for Molecular Motion. J. Comput. Phys. 229, 1077–1098.

Jernigan, J. B. & Kodaman, M. F. 2001 An Investigation of the Utility and Accuracy of the Table of Speed and Stopping Distances Specified in the Code of Virginia (Final Report). Virginia Transportation Research Council, Charlottesville, Virginia, 1–25.

Kapitza, S. P. 1996 The Phenomenological Theory of World Population Growth. Uspekhi Fizicheskikh Nauk 39, Russian Acad. Sci. 57–71.

Kermack, W. O. & McKendrick, A. G. 1927 Contributions to the Mathematical Theory of Epidemics. 1. Proceed. Royal Soc. 115A, 700–721.

Klimenko, A. Y. & Bilger, R. W. 1999 Conditional Moment Closure for Turbulent Combustion. Prog. Energ. Combust. 25, 595–687.

Kloeden, P. E. & Platen, E. 1992 Numerical Solution of Stochastic Differential Equations. Applications of Mathematics 23 (Stochastic Modeling and Applied Probability), Springer-Verlag, Berlin, Heidelberg, New York.

Kormondy, E. J. 1996 Concepts of Ecology. Prentice Hall, New York.

Kramers, H. A. 1940 Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions. Physica 7, 284–304.

Langevin, P. 1908 Sur la Théorie du Mouvement Brownien. Comptes Rendus Acad. Sci. Paris 146, 530–533.

Langhaar, H. L. 1951 Dimensional Analysis and Theory of Models. Wiley, New York.

L'Ecuyer, P. 1994 Uniform Random Number Generation. Ann. Oper. Res. 53, 77–120.

Li, Z. & Wang H. 2003 Drag Force, Diffusion Coefficient, and Electric Mobility of Small Particles. II. Application. Phys. Rev. E 68, 061207/1–13.

Liao, S. 2009 On the Reliability of Computed Chaotic Solutions of Non-Linear Differential Equations. Tellus 61A, 550–564.

Lorenz, E. N. 1963 Deterministic Nonperiodic Flow. J. Atmos. Sci. 20, 130–141.

Lorenz, E. N. 2006 Reflections on the Conception, Birth, and Childhood of Numerical Weather Prediction. Annu. Rev. Earth Pl. Sc. 34, 37–45.

Lubbers, J. & Graaff, R. 1998 A Simple and Accurate Formula for the Sound Velocity in Water. Ultrasound Med. Biol. 24, 1065–1068.

Luhar, A. K., Hibberd, M. F. & Hurley, P. J. 1996 Comparison of Closure Schemes Used to Specify the Velocity PDF in Lagrangian Stochastic Dispersion Models for Convective Conditions. Atmos. Environ. 30, 1407–1418.

Malthus, T. R. 1798 An Essay on the Principle of Population. J. Johnson, London.

May, R. M. 1975 Biological Populations Obeying Difference Equations: Stable Points, Stable Cycles, and Chaos. J. Theor. Biol. 51, 511–524.

May, R. M. (Editor) 1976 Theoretical Ecology: Principles and Applications. Blackwell, Oxford.

Meadows, D. H., Meadows, D. L., Randers, J. & Behrens III, W. W. 1972 The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind. Universe Books, New York.

Meadows, D. L., Behrens III, W. W., Meadows, D. H., Naill, R. F., Randers, J. & Zahn, E. K. O. 1974a Dynamics of Growth in a Finite World. Wright-Allen Press, MA.

Meadows, D. H., Meadows, D. L., Randers, J. & Behrens III, W. W. 1974b The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind. Second Edition, Universe Books, New York.

Meadows, D. H., Randers, J. & Meadows, D. L. 2004 Limits to Growth: The 30-Year Update. Chelsea Green Publ., White River Junction, VT.

Millikan, R. A. 1910 The Isolation of an Ion, a Precision Measurement of its Charge, and the Correction of Stokes' Law. Science 32, 436–448.

Millikan, R. A. 1917 A New Determination of e, N, and Related Constants. Philos. Mag. 34, 1–31.

Millikan, R. A. 1923 The General Law of Fall of a Small Spherical Body Through a Gas, and its Bearing Upon the Nature of Molecular Reflection From Surfaces. Phys. Rev. 22, 1–23.

Minier, J.-P. & Peirano, E. 2001 The PDF Approach to Turbulent Polydispersed Two-Phase Flows. Phys. Rep. 352, 1–214.

Moyal, J. E. 1949 Stochastic Processes and Statistical Physics. J. Roy. Stat. Soc. (London) B11, 150–210.

Murray, J. D. 2002 Mathematical Biology: I. An Introduction, Third Edition, Springer-Verlag, Berlin, Heidelberg.

Nagle, R. K., Saff, E. B. & Snider, A. D. 2008 Fundamentals of Differential Equations. Seventh Edition, Pearson Education, Boston, MA.

Newton, I. 1687 Philosophiæ Naturalis Principia Mathematica. Royal Society of London, UK.

Ortega, J. M. 1987 Matrix Theory: A Second Course. Plenum Press, New York.

Pawula, R. F. 1967 Approximation of Linear Boltzmann Equation by Fokker-Planck Equation. Phys. Rev. 162, 186–188.

Planck, M. 1917 Über Einen Satz der Statistischen Dynamik und Seine Erweiterung in der Quantentheorie. Sitzber. Preuß. Akad. Wiss., 324–341.

Pope, S. B. 1994 Lagrangian PDF Methods for Turbulent Flows. Annu. Rev. Fluid Mech. 26, 23–63.

Pope, S. B. 2000 Turbulent Flows. Cambridge University Press. Cambridge, UK.

Rainey, R. H. 1967 Natural Displacement of Pollution From the Great Lakes. Science 155, 1242–1243.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C. & Kaplan, A. 2003 Global Analyses of SST, Sea Ice and Night Marine Air Temperature Since the Late Nineteenth Century. J. Geophys. Res. 108, 4407–4443.

Risken, H. 1984 The Fokker-Planck Equation. Springer-Verlag, Berlin, Heidelberg, New York.

Roekaerts, D. 2002 Reacting Flows and Probability Density Function Methods. In: Closure Strategies for Turbulent and Transitional Flows, Edited by Launder, B. E. & Sandham, N. D., Cambridge Univ. Press, Cambridge, Chap. 10, 328–337.

Rössler, O. E. 1976 An Equation for Continuous Chaos. Phys. Lett. 57A, 397–398.

Ross, S. M. 2010 A First Course in Probability. Eighth Edition, Pearson Prentice Hall, Upper Saddle River, NJ.

Saltzman, B. 1962 Finite Amplitude Free Convection as an Initial Value Problem – I. J. Atmos. Sci. 19, 329–341.

Samuelson, P. A. 1939 Interaction Between the Multiplier Analysis and the Principle of Acceleration. Rev. Econ. Statistics 21, 75–78.

Sawford, B. L. 1991 Reynolds Number Effects in Lagrangian Stochastic Models of Turbulent Dispersion. Phys. Fluids A3, 1577–1586.

Scheiner, S. M. & Willig, M. R. 2005 Developing Unified Theories in Ecology as Exemplified with Diversity Gradients, Am. Nat. 166, 458–469.

Scheiner, S. M. & Willig, M. R. 2008 A General Theory of Ecology. Theor. Ecol. 1, 21–28.

Seinfeld, J. H. & Pandis, S. N. 2006 Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. Second Edition, Wiley, New York.

Shannon, C. H. 1948 A Mathematical Theory of Communication. Bell. System Tech. J. 27, 379–423.

Society of Fire Protection Engineers 1995 SFPE Handbook of Fire Protection Engineering. Second Edition, National Fire Protection Association, Edited by Beyler, C. L., Custer, R. L. P., Walton, W. D., Watts, J. M., Drysdale, D., Hall, J. R., Dinenno, P. J., Quincy, MA.

Sparrow, C. 1982 The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors. Applied Mathematical Sciences 41, Springer-Verlag, New York, Heidelberg, Berlin.

Stewart, J. 2006 Calculus: Concepts and Contexts. Third Edition, Brooks/Cole-Thomson, Belmont, CA.

Stokes, G. 1851 On the Effect of the Internal Friction of Fluids on the Motion of Pendulums. Transactions of the Cambridge Philosophical Society 9, 8–106.

Tans, P. 2008 Trends in Atmospheric Carbon Dioxide. NOAA/ESRL, Global Monitoring Division, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Earth System Research Laboratory, Boulder, CO, http://www.esrl.noaa.gov/gmd/ccgg/trends/.

Turchin, P. 2001 Does Population Ecology Have General Laws? Oikos 94, 17–26.

Turchin, P. 2003 Complex Population Dynamics: A Theoretical/Empirical Synthesis. Princeton Univ. Press, Princeton and Oxford.

Turner, G. M. 2008 A Comparison of The Limits to Growth With 30 Years of Reality. Global Environ. Chang. 18, 397– 411.

Tyagi, M., Jenny, P., Lunati, I. & Tchelepi, H. A. 2008 A Lagrangian, Stochastic Modeling Framework for Multi-Phase Flow in Porous Media. J. Comput. Phys. 227, 6696–6714.

U.S. Dept. of Energy 2008 Annual Energy Report 2007, Washington DC.

Verhulst, P. F. 1838 Notice sur la Loi que la Population Poursuit dans son Accroisement. Correspondence Mathématique et Physique 10, 113–121.

Wiggins, S. 2010 Introduction to Applied Nonlinear Dynamical Systems and Chaos. Texts in Applied Mathematics, Second Edition, Springer-Verlag, New York.

World Almanac Books 2010 The World Almanac and Book of Facts 2010. World Almanac Books, New York.

XFOIL (http://web.mit.edu/drela/Public/web /xfoil/).

Yafia, R. 2009 Dynamics and Numerical Simulations in a Production and Development of Red Blood Cells Model with one Delay. Commun. Nonlinear Sci. 14, 582–592.

Yao, L.-S. 2007 Is a Direct Numerical Simulation of Chaos Possible? A Study of a Model Nonlinearity. Int. J. Heat Mass Tran. 50, 2200–2207.

Yao, L.-S & Hughes, D. 2008 Comment on "Computational Periodicity as Observed in a Simple System" by Edward N. Lorenz (2006). Tellus 60A, 803–805.

Yao, L.-S. 2010 Computed Chaos or Numerical Errors. Nonlinear Anal. Modell. Control 15, 109–126.

# Author Index

# Subject Index

**Boldface** page numbers indicate principal references.

# About the Author

Dr. Stefan Heinz is a Professor of Mathematics at the University of Wyoming. He holds a Ph.D. in Physics from the Heinrich-Hertz Institute, Berlin. His research interests are in mathematical modeling, multiscale processes, stochastic analysis, Monte Carlo simulations, computational fluid dynamics, turbulence, combustion, and multiphase flows. He has authored more than seventy refereed publications and the textbook Statistical Mechanics of Turbulent Flows (Springer, 2003). For more than ten years he has taught a variety of courses: calculus, probability, ordinary, partial, and stochastic differential equations, applied mathematics, and deterministic and stochastic mathematical modeling. His exceptional teaching was awarded in 2007 by the College of Arts and Sciences Extraordinary Merit in Teaching Award. In 2008 he was honored as Adjunct Professor of Mechanical Engineering. He has held visiting professor appointments at ETH Zurich, Delft Technical University, and the National Center for Atmospheric Research (NCAR) at Boulder.