Kurt Marti

Stochastic Optimization Methods

Applications in Engineering and Operations Research

Fourth Edition

Springer

Stochastic Optimization Methods

Kurt Marti

Stochastic Optimization Methods

Applications in Engineering and Operations Research

Fourth Edition

Springer

Kurt Marti
Institute for Mathematics and Computer
Science
Federal Armed Forces University Munich
Munich, Germany

ISBN 978-3-031-40058-2 ISBN 978-3-031-40059-9 (eBook)
https://doi.org/10.1007/978-3-031-40059-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

Optimization problems in practice depend mostly on several model parameters, noise factors, uncontrollable parameters, etc., which are not given fixed quantities at the planning stage. Due to several types of stochastic uncertainties (physical uncertainty, economic uncertainty, statistical uncertainty, model uncertainty), these parameters must be modeled by random variables having a certain probability distribution. In most cases at least certain moments of this distribution are known.

In order to cope with these uncertainties, a basic procedure in the behavior of the structure/system from the prescribed performance (output, behavior), i.e., the *tracking error*, is compensated by (online) input corrections. However, the online correction of a system/structure is often time consuming and causes mostly increasing expenses (correction, repair, or recourse costs). Very large recourse costs may arise in case of damages or failures of the plant. This can be omitted to a large extent by taking into account already at the planning stage the possible consequences of the tracking errors and the known prior and sample information about the random data of the problem. Hence, instead of relying on ordinary deterministic parameter optimization methods—based on some nominal parameter values—and applying then just some correction actions, stochastic optimization methods should be applied: Incorporating the consequences of stochastic parameter variations into the optimization process, large and increasing recourse, repair, recovery costs can be omitted or at least reduced to a large extent.

Consequently, for the computation of robust optimal decisions/designs, i.e., optimal decisions which are insensitive with respect to random parameter variations, appropriate deterministic substitute problems must be formulated first. Based on decision theoretical principles, these substitute problems depend on probabilities of failure/success and/or on more general expected cost/loss terms. Since probabilities and expectations are defined by multiple integrals in general, the resulting often nonlinear and also non-convex deterministic substitute problems can be solved by approximate methods only. Two basic types of deterministic substitute problems occur mostly in practice:

- *Minimization of the expected primary costs subject to expected recourse cost constraints (reliability constraints) and remaining deterministic constraints, e.g., box constraints.*
- *Expected Total Cost Minimization Problems subject to deterministic constraints.*

In case of piecewise constant cost functions, probabilistic objective functions and/or probabilistic constraints occur.

Main analytical properties of the substitute problems have been examined in the first three editions of the book, where also appropriate deterministic and stochastic approximation and solution procedures can be found.

The aim of the present fourth edition is the presentation of updated methods for the transformation of actual technical and economic optimization problems with random parameters into appropriate deterministic substitute problems. Hence, updated analytical and numerical tools are provided for the approximate computation of robust optimal decisions/designs/control, as needed in concrete engineering/economic applications.

Last but not least I would like to thank Dipl. Math. Ina Stein, Munich, for her excellent support in the LaTeX-typesetting as well as in the final proofreading. Moreover, I am indebted to Springer Nature for inviting a new edition of the monograph *Stochastic Optimization Methods*. I would like to thank especially the Senior Editor for Business/Economics/Operations Research of Springer-Verlag Heidelberg, Germany, Christian Rauscher and the Springer Editors Yvonne Schwark-Reiber, Books Editorial Projects Management, and Jialin Yan, Book Editor Operations Research and Management, Information Systems and Applied Statistics, for their advice during the preparation of this new edition.

Munich, Germany Kurt Marti
March 2024

Contents

Chapter 1
Stochastic Optimization Methods

Abstract Basic methods for treating stochastic optimization problems (SOP), hence, optimization problems with random data are presented: Optimization problems in practice depend mostly on several model parameters, noise factors, uncontrollable parameters, etc., which are not given fixed quantities at the planning stage. Typical examples from engineering and economics/operations research are: Material parameters (e.g., elasticity moduli, yield stresses, allowable stresses, moment capacities, specific gravity), external loadings, friction coefficients, moments of inertia, length of links, mass of links, location of the center of gravity of links, manufacturing errors, tolerances, noise terms, demand parameters, technological coefficients in input-output functions, cost factors, interest rates, exchange rates, etc. Due to several types of stochastic uncertainties (physical uncertainty, economic uncertainty, statistical uncertainty, model uncertainty) these parameters must be modeled by random variables having a certain probability distribution. In most cases at least certain moments of this distribution are known.n order to cope with these uncertainties, a basic procedure in the engineering/economic practice is to replace first the unknown parameters by some chosen nominal values, e.g., estimates, guesses, of the parameters. Then, the resulting and mostly increasing deviation of the performance (output, behavior) of the structure/system from the prescribed performance (output, behavior), i.e., the *tracking error*, is compensated by (online) input corrections. However, the online correction of a system/structure is often time consuming and causes mostly increasing expenses (correction or recourse costs). Very large recourse costs may arise in case of damages or failures of the plant. This can be omitted to a large extent by taking into account already at the planning stage the possible consequences of the tracking errors and the known prior and sample information about the random data of the problem. Hence, instead of relying on ordinary deterministic parameter optimization methods - based on some nominal parameter values—and applying then just some correction actions, stochastic optimization methods should be applied: Incorporating stochastic parameter variations into the optimization process, expensive and increasing online correction expenses can be omitted or at least reduced to a large extent. Consequently, for the computation of robust optimal decisions/designs, i.e., optimal decisions which are insensitive with respect to random parameter variations, appropriate deterministic substitute problems must be formulated first. Based on decision theoretical principles, these substitute problems depend on probabilities

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9_1

of failure/success and/or on more general expected cost/loss terms. Two basic types of deterministic substitute problems occur mostly in practice:

- *Reliability-Based Optimization Problems*: primary cost minimization subject to expected recourse (correction) cost constraints: Minimization of the expected primary costs subject to expected recourse cost constraints (reliability constraints) and remaining deterministic constraints, e.g., box constraints. In case of piecewise constant cost functions, probabilistic objective functions and/or probabilistic constraints occur;
- *Expected Total Cost Minimization Problems*: Minimization of the expected total costs (costs of construction, design, recourse/correction, repair costs, etc.) subject to the remaining deterministic constraints.

Since probabilities and expectations are defined by multiple integrals in general, the resulting often nonlinear and also non-convex deterministic substitute problems can be solved by approximate methods only.

1.1 Introduction

Many concrete problems from engineering, economics, operations research, etc., can be formulated by an optimization problem of the type

$$\begin{aligned} \min \quad & f_0(a,x) && && (1.1a)\\ \text{s.t.} \quad & && && \\ & f_i(a,x) \le 0, && i = 1,\dots,m_f && (1.1b)\\ & g_i(a,x) = 0, && i = 1,\dots,m_g && (1.1c)\\ & x \in D_0. && && (1.1d) \end{aligned}$$

Here, the objective (goal) function $f_0 = f_0(a,x)$ and the constraint functions $f_i = f_i(a,x), i = 1,\dots,m_f$ and $g_i = g_i(a,x), i = 1,\dots,m_g$, defined on a joint subset of $\mathbb{R}^\nu \times \mathbb{R}^r$, depend on a decision, design, control orinput vector $x = (x_1, x_2, \dots, x_r)^T$ and a vector $a = (a_1, a_2, \dots, a_\nu)^T$ ofmodel parameters. Typical model parameters in technical applications, operations research, and economics are material parameters, external load parameters, cost factors, technological parameters in input-output operators, demand factors. Furthermore, manufacturing and modeling errors, disturbances or noise factors, etc., may occur. Frequent decision, control, or input variables are material, topological, geometrical and cross-sectional design variables in structural optimization [23], forces and moments in optimal control of dynamic systems and factors of production in operations research and economic design.

The objective function (1.1a) to be optimized describes the aim, the goal of the modeled optimal decision/design problem or the performance of a technical, economic system or process to be controlled optimally. Furthermore, the constraints

(1.1b)–(1.1d) represent the operating conditions guaranteeing a safe structure, a correct functioning of the underlying system, process, etc. Note that the constraint (1.1d) with a given, fixed convex subset $D_0 \subset \mathbb{R}^r$ summarizes all (deterministic) constraints being independent of unknown model parameters a, as, e.g., box constraints:

$$x^L \leq x \leq x^U \tag{1.1e}$$

with given bounds x^L, x^U.

Important concrete optimization problems, which may be formulated, at least approximate, this way, are problems from optimal design of mechanical structures and structural systems [1, 23, 43, 48], adaptive trajectory planning for robots [2, 3, 14, 30, 37, 45], adaptive control of dynamic system [46, 47], optimal design of economic systems [22], production planning, manufacturing [26, 38] and sequential decision processes [34], etc.

In *optimal control*, cf. Chap. 3, the input vector $x := u(\cdot)$ is interpreted as a function, a *control or input function* $u = u(t), t_0 \leq t \leq t_f$, on a certain given time interval $[t_0, t_f]$. Moreover, see Chap. 3, the objective function $f_0 = f_0(a, u(\cdot))$ is defined by a certain integral over the time interval $[t_0, t_f]$. In addition, the constraint functions $f_j = f_j(a, u(\cdot))$ are defined by integrals over $[t_0, t_f]$, or $f_j = f_j(t, a, u(t))$ may be functions of time t and the control input $u(t)$ at time t.

A basic problem in practice is that the vector of model parameters $a = (a_1, \ldots, a_\nu)^T$ is not a given, fixed quantity. Model parameters are often unknown, only partly known and/or may vary randomly to some extent.

Several techniques have been developed in the recent years in order to cope with uncertainty with respect to model parameters a. A well-known basic method, often used in engineering practice, is the following two-step procedure [3, 14, 37, 45, 46]:

(I) **Parameter Estimation and Approximation**:
First, replace first the ν-vector a of the unknown or stochastic varying model parameters $a_1, \ldots, a_\nu$ by some estimated/chosen fixed vector a_0 of so-called *nominal* values $a_{0l}, l = 1, \ldots, \nu$.
Then, apply an optimal decision (control) $x^* = x^*(a_0)$ with respect to the resulting approximate optimization problem

$$\begin{aligned}
&\min \quad f_0(a_0, x) && && (1.2a)\\
&\text{s.t.} \\
&\qquad f_i(a_0, x) \leq 0, && i = 1, \ldots, m_f && (1.2b)\\
&\qquad g_i(a_0, x) = 0, && i = 1, \ldots, m_g && (1.2c)\\
&\qquad x \in D_0. && && (1.2d)
\end{aligned}$$

Due to the deviation of the actual parameter vector a from the nominal vector a_0 of model parameters, deviations of the actual state, trajectory or performance of the system from the prescribed state, trajectory, goal values occur.

(II) **Compensation or correction**:
Then, the deviation of the actual state, trajectory or performance of the system from the prescribed values/functions is compensated by online measurement and correction actions (decisions or controls). Consequently, in general, increasing measurement and correction expenses result in course of time.

Considerable improvements of this standard procedure can be obtained by taking into account already at the planning stage, i.e., offline, the mostly available a priori (e.g., the type of random variability) and sample information about the parameter vector a. Indeed, based, e.g., on some structural insight, or by parameter identification methods, regression techniques, calibration methods, etc., in most cases information about the vector a of model parameters can be extracted. Repeating this information gathering procedure at some later time points $t_j > t_0$ (= initial time point), $j = 1, 2, \ldots$, adaptive decision/control procedures occur [34].

Based on the inherent random nature of the parameter vector a, the observation or measurement mechanism, resp., or adopting a Bayesian approach concerning unknown parameter values [6], here we make the following basic assumption:

Stochastic (Probabilistic) Uncertainty : The unknown parameter vector a is a realization

$$a = a(\omega)\omega \in \Omega, \tag{1.3}$$

of a random ν-vector $a(\omega)$ on a certain probability space $(\Omega, \mathcal{A}_0, P)$, where the probability distribution $P_{a(\cdot)}$ of $a(\omega)$ is known, or it is known that $P_{a(\cdot)}$ lies within a given range W of probability measures on $\mathbb{R}^\nu$. Using a Bayesian approach, the probability distribution $P_{a(\cdot)}$ of $a(\omega)$ may also describe the subjective or personal probability of the decision maker, the designer.

Hence, in order to take into account the stochastic variations of the parameter vector a, to incorporate the a priori and/or sample information about the unknown vector a, resp., the standard approach "insert a certain nominal parameter vector a_0, and correct then the resulting error", must be replaced by a more appropriate deterministic substitute problem for the basic optimization problem (1.1a)–(1.1d) under stochastic uncertainty.

1.2 Deterministic Substitute Problems: Basic Formulation

The proper selection of a deterministic substitute problem is a decision theoretical task, see [27]. Hence, for (1.1a)–(1.1d) we have first to consider the *outcome map*

$$\begin{aligned} e &= e(a, x) \\ &:= \Big(f_0(a, x),\, f_1(a, x), \ldots, f_{m_f}(a, x),\, g_1(a, x), \ldots, g_{m_g}(a, x) \Big)^T, \\ a &\in \mathbb{R}^\nu,\, x \in \mathbb{R}^r,\, (x \in D_0), \end{aligned} \tag{1.4a}$$

and to evaluate then the outcomes $e \in \mathcal{E} \subset \mathbb{R}^{1+m_0}$, $m_0 := m_f + m_g$, by means of certain loss or cost functions

$$\gamma_i : \mathcal{E} \to \mathbb{R}, \qquad i = 0, 1, \ldots, m \tag{1.4b}$$

with an integer $m \geq 0$. For the processing of the numerical outcomes $\gamma_i\big(e(a, x)\big)$, $i = 0, 1, \ldots, m$, there are two basic concepts:

1.2.1 Minimum or Bounded Expected Costs

Consider the vector of (conditional) expected losses or costs

$$\mathbf{F}(x) = \begin{pmatrix} F_0(x) \\ F_1(x) \\ \vdots \\ F_m(x) \end{pmatrix} := \begin{pmatrix} E\gamma_0(e(a(\omega), x)) \\ E\gamma_1(e(a(\omega), x)) \\ \vdots \\ E\gamma_m(e(a(\omega), x)) \end{pmatrix}, \qquad x \in \mathbb{R}^r, \tag{1.5}$$

where the (conditional) expectation "E" is taken with respect to the time history $\mathfrak{A} = \mathfrak{A}_t$, $(\mathfrak{A}_j) \subset \mathfrak{A}_0$ up to a certain time point t or stage j. A short definition of expectations is given in Sect. 1.3, for more details, see, e.g., [5, 18, 40].

Having different expected cost or performance functions $F_0, F_1, \ldots, F_m$ to be minimized or bounded, as a basic deterministic substitute problem for (1.1a)–(1.1d) with a random parameter vector $a = a(\omega)$ we may consider the multi-objective expected cost minimization problem

$$\text{"min"} \quad \mathbf{F}(x) \tag{1.6a}$$

$$\text{s.t.} \quad x \in D_0. \tag{1.6b}$$

Obviously, a good compromise solution x^* of this vector optimization problem should have at least one of the following properties [13, 41]:

Definition 1.1

(a) A vector $x^0 \in D_0$ is called a **functional-efficient** or **Pareto optimal** solution of the vector optimization problem (1.6a), (1.6b) if there is no $x \in D_0$ such that

$$F_i(x) \leq F_i(x^0), \qquad i = 0, 1, \ldots, m \tag{1.7a}$$

and

$$F_{i_0}(x) < F_{i_0}(x^0) \qquad \text{for at least one } i_0,\ 0 \leq i_0 \leq m. \tag{1.7b}$$

(b) A vector $x^0 \in D_0$ is called a **weak functional-efficient** or **weak Pareto optimal** solution of (1.6a)–(1.6b) if there is no $x \in D_0$ such that

$$F_i(x) < F_i(x^0), \quad i = 0, 1, \ldots, m \tag{1.8}$$

(Weak) Pareto optimal solutions of (1.6a)–(1.6b) may be obtained now by means of scalarizations of the vector optimization problem (1.6a)–(1.6b). Three main versions are stated in the following:

(I) *Minimization of primary expected cost/loss under expected cost constraints*

$$\begin{aligned} \min \quad & F_0(x) && \text{(1.9a)} \\ \text{s.t.} \quad & \\ & F_i(x) \le F_i^{\max}, \quad i = 1, \ldots, m && \text{(1.9b)} \\ & x \in D_0. && \text{(1.9c)} \end{aligned}$$

Here, $F_0 = F_0(x)$ is assumed to describe the primary goal of the design/decision-making problem, while $F_i = F_i(x)$, $i = 1, \ldots, m$, describe secondary goals. Moreover, $F_i^{\max}$, $i = 1, \ldots, m$, denote given upper cost/loss bounds.

Remark 1.1 An optimal solution x^* of (1.9a)–(1.9c) is a weak Pareto optimal solution of (1.6a)–(1.6b).

(II) *Minimization of the total weighted expected costs*
Selecting certain positive weight factors $c_0, c_1, \ldots, c_m$, the expected weighted total costs are defined by

$$\tilde{F}(x) := \sum_{i=0}^{m} c_i F_i(x) = Ef\Big(a(\omega), x\Big), \tag{1.10a}$$

where

$$f(a, x) := \sum_{i=0}^{m} c_i \gamma_i\Big(e(a, x)\Big). \tag{1.10b}$$

Consequently, minimizing the expected weighted total costs $\tilde{F} = \tilde{F}(x)$ subject to the remaining deterministic constraint (1.1d), the following deterministic substitute problem for (1.1a)–(1.1d) occurs

$$\begin{aligned} \min \quad & \sum_{i=0}^{m} c_i F_i(x) && \text{(1.11a)} \\ \text{s.t.} \quad & x \in D_0. && \text{(1.11b)} \end{aligned}$$

Remark 1.2 Let $c_i > 0, i = 1, 1, \ldots, m$, be any positive weight factors. Then, an optimal solution x^* of (1.11a)–(1.11b) is a Pareto optimal solution of (1.6a)–(1.6b).

(III) *Minimization of the maximum weighted expected costs*
Instead of adding weighted expected costs, we may consider the maximum of the weighted expected costs:

$$\tilde{F}(x) := \max_{0\le i\le m} c_i F_i(x) = \max_{0\le i\le m} c_i E\gamma_i\Big(e\big(a(\omega), x\big)\Big). \tag{1.12}$$

Here again, $c_0, c_1, \ldots, c_m$, are positive weight factors.
Thus, minimizing $\tilde{F} = \tilde{F}(x)$ we have the deterministic substitute problem

$$\min \quad \max_{0\le i\le m} c_i F_i(x) \tag{1.13a}$$

$$\text{s.t.} \quad x \in D_0. \tag{1.13b}$$

Remark 1.3 Let $c_i, i = 0, 1, \ldots, m$, be any positive weight factors. An optimal solution of x^* of (1.13a)–(1.13b) is a weak Pareto optimal solution of (1.6a)–(1.6b).

1.2.2 Minimum or Bounded Maximum Costs (Worst Case)

Instead of taking expectations, we may consider the worst case with respect to the cost variations caused by the random parameter vector $a = a(\omega)$. Hence, the random cost function

$$\omega \to \gamma_i\Big(e\big(a(\omega), x\big)\Big) \tag{1.14a}$$

is evaluated by means of

$$F_i^{\sup}(x) := \operatorname{ess\,sup} \gamma_i\Big(e\big(a(\omega), x\big)\Big), \quad i = 0, 1, \ldots, m. \tag{1.14b}$$

Here, ess sup (...) denotes the (conditional) essential supremum with respect to the random vector $a = a(\omega)$, given information $\mathfrak{A}$, i.e., the infimum of the supremum of (1.14a) on sets $A \in \mathfrak{A}_0$ of (conditional) probability one, see, e.g., [40].

Consequently, the vector function $\mathbf{F} = \mathbf{F}^{\sup}(x)$ is then defined by

$$\mathbf{F}^{\sup}(x) = \begin{pmatrix} F_0(x) \\ F_1(x) \\ \vdots \\ F_m(x) \end{pmatrix} := \begin{pmatrix} \operatorname{ess\,sup} \gamma_0\Big(e\big(a(\omega), x\big)\Big) \\ \operatorname{ess\,sup} \gamma_1\Big(e\big(a(\omega), x\big)\Big) \\ \vdots \\ \operatorname{ess\,sup} \gamma_m\Big(e\big(a(\omega), x\big)\Big) \end{pmatrix}. \tag{1.15}$$

Working with the vector function $\mathbf{F} = \mathbf{F}^{\sup}(x)$, we have then the vector minimization problem

$$\text{“min”} \quad \mathbf{F}^{\sup}(x) \tag{1.16a}$$

$$\text{s.t.} \quad x \in D_0. \tag{1.16b}$$

By scalarization of (1.16a)–(1.16b) we then obtain deterministic substitute problems for (1.1a)–(1.1d) related to the substitute problem (1.6a)–(1.6b) introduced in Sect. 1.2.1.

More details on the selection and solution of appropriate deterministic substitute problems for (1.1a)–(1.1d) are given in the next sections. Deterministic substitute problems for optimal control problems under stochastic uncertainty are considered in Chap. 3.

1.3 Optimal Decision/Design Problems with Random Parameters

In the optimal design of technical or economic structures/systems, in optimal decision problems arising in technical or economic systems, resp., two basic classes of criteria appear.

First there is a primary cost function

$$G_0 = G_0(a, x). \tag{1.17a}$$

Important examples are the total weight or volume of a mechanical structure, the costs of construction, design of a certain technical or economic structure/system, or the negative utility or reward in a general decision situation. Basic examples in optimal control, cf. Chap. 3, are the total run time, the total energy consumption of the process or a weighted mean of these two cost functions.

For the representation of the structural/system safety or failure, for the representation of the admissibility of the state, or for the formulation of the basic operating conditions of the , certain **state, performance or response functions**

$$y_i = y_i(a, x), \ i = 1, \ldots, m_y \tag{1.17b}$$

are chosen. In structural design these functions are also called “limit state functions” or “safety margins”. Frequent examples are some displacement, stress, load (force and moment) components in structural design, or more general system output functions in engineering design. Furthermore, production functions and several cost functions are possible performance functions in production planning problems, optimal mix problems, transportation problems, allocation problems and other problems of economic decision.

In (1.17a,b), the design or input vector x denotes the r-vector of design or input variables, $x_1, x_2, \ldots, x_r$, as, e.g., structural dimensions, sizing variables, such as cross-sectional areas, thickness in structural design, or factors of production, actions in economic decision problems. For the decision, design or input vector x one has mostly some basic deterministic constraints, e.g., nonnegativity constraints, box constraints, represented by

$$x \in D, \tag{1.17c}$$

where D is a given convex subset of $\mathbb{R}^r$. Moreover, a is the ν-vector of model parameters. In optimal structural/engineering design

$$a = \begin{pmatrix} p \\ R \end{pmatrix} \tag{1.17d}$$

is composed of the following two subvectors: R is the m-vector of the acting external loads or structural/system inputs, e.g., wave, wind loads, payload, etc. Moreover, p denotes the $(\nu - m)$-vector of the further model parameters, as, e.g., material parameters, like strength parameters, yield/allowable stresses, elastic moduli, plastic capacities, etc., of the members of a mechanical structure, parameters of an electric circuit, such as resistances, inductances, capacitances, the manufacturing tolerances and weight or more general cost coefficients.

In linear programming, as, e.g., in production planning problems,

$$a = (A, b, c) \tag{1.17e}$$

is composed of the $m \times r$ matrix A of technological coefficients, the demand m-vector b and the r-vector c of unit costs.

Based on the m_y-vector of state functions

$$y(a, x) := \Big(y_1(a, x), y_2(a, x), \ldots, y_{m_y}(a, x)\Big)^T, \tag{1.17f}$$

the admissible or safe states of the structure/system can be characterized by the condition

$$y(a, x) \in B, \tag{1.17g}$$

where B is a certain subset of $\mathbb{R}^{m_y}$; $B = B(a)$ may depend also on some model parameters.

In production planning problems, typical operating conditions are given, cf. (1.17e), by

$$y(a, x) := Ax - b \geq 0 \qquad \text{or} \qquad y(a, x) = 0,\ x \geq 0. \tag{1.18a}$$

In mechanical structures/structural systems, the safety (survival) of the structure/system is described by the operating conditions

$$y_i(a, x) > 0 \quad \text{for all} \quad i = 1, \ldots, m_y \tag{1.18b}$$

with state functions $y_i = y_i(a, x),\ i = 1, \ldots, m_y$, depending on certain response components of the structure/system, such as displacement, stress, force, moment components.

Hence, a failure occurs if and only if the structure/system is in the i-th failure mode (failure domain)

$$y_i(a, x) \leq 0 \tag{1.18c}$$

for at least one index i, $1 \leq i \leq m_y$.

Note 1.1 The number m_y of safety margins or limit state functions $y_i = y_i(a, x),\ i = 1, \ldots, m_y$, may be very large. For example, in optimal plastic design the limit state functions are determined by the extreme points of the admissible domain of the dual pair of static/kinematic LPs related to the equilibrium and linearized convex yield condition, see [32, 33].

Basic problems in optimal decision/design are

(I) *Primary (construction, planning, investment, etc.) cost minimization under operating or safety conditions*

$$\begin{aligned} &\min \quad G_0(a, x) && (1.19a)\\ &\text{s.t.} \\ &\qquad y(a, x) \in B && (1.19b)\\ &\qquad x \in D. && (1.19c) \end{aligned}$$

Obviously we have $B = (0, +\infty)^{m_y}$ in (1.18b) and $B = [0, +\infty)^{m_y}$ or $B = \{0\}$ in (1.18a).

(II) *Failure or recourse cost minimization under primary cost constraints*

$$\begin{aligned} &\text{“min”} \quad \gamma\Big(y(a, x)\Big) && (1.20a)\\ &\text{s.t.} \\ &\qquad G_0(a, x) \leq G^{\max} && (1.20b)\\ &\qquad x \in D. && (1.20c) \end{aligned}$$

In (1.20a) $\gamma = \gamma(y)$ is a scalar or vector valued cost/loss function evaluating violations of the operating conditions (1.19b). Depending on the application, these costs are called “failure” or “recourse” costs [20, 21, 31, 39, 43, 44]. As already discussed in Sect. 1.1, solving problems of the above type, a basic difficulty is the uncertainty about the true value of the vector a of model parameters or the (random) variability of a. In practice, due to several types of uncertainties such as, see [49],

- physical uncertainty (variability of physical quantities, like material, loads, dimensions, etc.)
- economic uncertainty (trade, demand, costs, etc.)
- statistical uncertainty (e.g., estimation errors of parameters due to limited sample data)
- model uncertainty (model errors).

The ν-vector a of model parameters must be modeled by a random vector

$$a = a(\omega), \omega \in \Omega, \tag{1.21a}$$

on a certain probability space $(\Omega, \mathfrak{A}_0, P)$ with sample space Ω having elements ω, see (1.3). For the mathematical representation of the corresponding (conditional) probability distribution $P_{a(\cdot)} = P^{\mathfrak{A}}_{a(\cdot)}$ of the random vector $a = a(\omega)$ (given the time history or information $\mathfrak{A} \subset \mathfrak{A}_0$), two main distribution models are taken into account in practice:

(i) Discrete probability distributions,
(ii) Continuous probability distributions.

In the first case there is a finite or countably infinite number $l_0 \in \mathbb{N} \cup \{\infty\}$ of realizations or scenarios $a^l \in \mathbb{R}^\nu, \quad l = 1, \ldots, l_0$,

$$P\Big(a(\omega) = a^l\Big) = \alpha_l, \quad l = 1, \ldots, l_0, \tag{1.21b}$$

taken with probabilities $\alpha_l, \ l = 1, \ldots, l_0$.

In the second case, the probability that the realization $a(\omega) = a$ lies in a certain (measurable) subset $B \subset \mathbb{R}^\nu$ is described by the multiple integral

$$P\Big(a(\omega) \in B\Big) = \int\limits_B \varphi(a)\, da \tag{1.21c}$$

with a certain probability density function $\varphi = \varphi(a) \geq 0, \ a \in \mathbb{R}^\nu, \ \int \varphi(a) da = 1$.

The properties of the probability distribution $P_{a(\cdot)}$ may be described—fully or in part—by certain numerical characteristics, called parameters of $P_{a(\cdot)}$. These distribution parameters $\theta = \theta_h$ are obtained by considering expectations

$$\theta_h := Eh\Big(a(\omega)\Big) \tag{1.22a}$$

of some (measurable) functions

$$(h \circ a)(\omega) := h\Big(a(\omega)\Big) \tag{1.22b}$$

composed of the random vector $a = a(\omega)$ with certain (measurable) mappings

$$h : \mathbb{R}^{\nu} \longrightarrow \mathbb{R}^{s_h}, \quad s_h \geq 1. \tag{1.22c}$$

According to the type of the probability distribution $P_{a(\cdot)}$ of $a = a(\omega)$, the expectation $Eh\Big(a(\omega)\Big)$ is defined, cf. [4, 5], by

$$Eh\Big(a(\omega)\Big) = \begin{cases} \sum\limits_{l=1}^{l_0} h\Big(a^l\Big)\, \alpha_l, & \text{in the discrete case (1.21b)} \\ \int\limits_{\mathbb{R}^{\nu}} h(a)\varphi(a)\, da, & \text{in the continuous case (1.21c).} \end{cases} \tag{1.22d}$$

Further distribution parameters θ are functions

$$\theta = \Psi(\theta_{h_1}, \ldots, \theta_{h_s}) \tag{1.23}$$

of certain "h-moments" $\theta_{h_1}, \ldots, \theta_{h_s}$ of the type (1.22a). Important examples of the type (1.22a), (1.23), resp., are the expectation

$$\overline{a} = Ea(\omega) \quad (\text{for } h_1(a) := \overline{a},\ \overline{a} \in \mathbb{R}^{\nu}) \tag{1.24a}$$

and the covariance matrix

$$Q := E\Big(a(\omega) - \overline{a}\Big)\Big(a(\omega) - \overline{a}\Big)^T = Ea(\omega)a(\omega)^T - \overline{a}\,\overline{a}^T \tag{1.24b}$$

of the random vector $a = a(\omega)$.

Due to the stochastic variability of the random vector $a(\cdot)$ of model parameters, and since the realization $a(\omega) = a$ is not available at the decision-making stage, the optimal design problem (1.19a)–(1.19c) or (1.20a)–(1.20c) under stochastic uncertainty cannot be solved directly.

Hence, appropriate deterministic substitute problems must be chosen taking into account the randomness of $a = a(\omega)$, cf. Sect. 1.2.

1.4 Deterministic Substitute Problems in Optimal Decision/Design

According to Sect. 1.2, a basic deterministic substitute problem in optimal design under stochastic uncertainty is the minimization of the total expected costs including the expected costs of failure

$$\min \quad c_G \cdot EG_0\Big(a(\omega), x\Big) + c_f \cdot p_f(x) \tag{1.25a}$$

$$\text{s.t.} \quad x \in D. \tag{1.25b}$$

Here,

$$p_f = p_f(x) := P\left(y\Big(a(\omega), x\Big) \notin B\right) \tag{1.25c}$$

is the probability of failure or the probability that a safe function of the structure, the system is not guaranteed. Furthermore, c_G is a certain weight factor, and $c_f > 0$ describes the failure or recourse costs. In the present definition of expected failure costs, constant costs for each realization $a = a(\omega)$ of $a(\cdot)$ are assumed. Obviously, it is

$$p_f(x) = 1 - p_s(x) \tag{1.25d}$$

with the probability of safety or survival

$$p_s(x) := P\left(y\Big(a(\omega), x\Big) \in B\right). \tag{1.25e}$$

In case (1.18b) we have

$$p_f(x) = P\left(y_i\Big(a(\omega), x\Big) \le 0 \text{ for at least one index } i,\ 1 \le i \le m_y\right). \tag{1.25f}$$

The objective function (1.25a) may be interpreted as the Lagrangian (with given cost multiplier c_f) of the following reliability-based optimization (RBO) problem, cf. [1, 29, 39, 43, 49]:

$$\min \quad EG_0\Big(a(\omega), x\Big) \tag{1.26a}$$

$$\text{s.t.}$$

$$p_f(x) \le \alpha^{\max} \tag{1.26b}$$

$$x \in D, \tag{1.26c}$$

where $\alpha^{\max} > 0$ is a prescribed maximum failure probability, e.g., $\alpha^{\max} = 0.001$, cf. (1.19a)–(1.19c).

The "dual" version of (1.26a)–(1.26c) reads

$$\min \quad p_f(x) \tag{1.27a}$$

$$\text{s.t.}$$

$$EG_0\Big(a(\omega), x\Big) \le G^{\max} \tag{1.27b}$$

$$x \in D \tag{1.27c}$$

with a maximal (upper) cost bound $G^{\max}$, see (1.20a)–(1.20c).

1.4.1 Expected Cost or Loss Functions

Further substitute problems are obtained by considering more general expected failure or recourse cost functions

$$\Gamma(x) = E\gamma\left(y\Big(a(\omega), x\Big)\right) \tag{1.28a}$$

arising from structural systems weakness or failure, or because of false operation. Here,

$$y\Big(a(\omega), x\Big) := \left(y_1\Big(a(\omega), x\Big), \ldots, y_{m_y}\Big(a(\omega), x\Big)\right)^T \tag{1.28b}$$

is again the random vector of state or performance functions, and

$$\gamma : \mathbb{R}^{m_y} \to \mathbb{R}^{m_\gamma} \tag{1.28c}$$

is a scalar or vector valued cost or loss function. In case $B = (0, +\infty)^{m_y}$ or $B = [0, +\infty)^{m_y}$ it is often assumed that $\gamma = \gamma(y)$ is a non-increasing function, hence,

$$\gamma(y) \geq \gamma(z), \quad \text{if} \quad y \leq z, \tag{1.28d}$$

where inequalities between vectors are defined component-by-component.

Example 1.1 If $\gamma(y) = 1$ for $y \in B^c$ (complement of B) and $\gamma(y) = 0$ for $y \in B$, then $\Gamma(x) = p_f(x)$.

Example 1.2 Suppose that $\gamma = \gamma(y)$ is a nonnegative measurable scalar function on $\mathbb{R}^{m_y}$ such that

$$\gamma(y) \geq \gamma_0 > 0 \text{ for all } y \notin B \tag{1.29a}$$

with a constant $\gamma_0 > 0$. Then for the probability of failure we find the following upper bound

$$p_f(x) = P\left(y\Big(a(\omega), x\Big) \notin B\right) \leq \frac{1}{\gamma_0} E\gamma\left(y\Big(a(\omega), x\Big)\right), \tag{1.29b}$$

where the right-hand side of (1.29b) is obviously an expected cost function of type (1.28a)–(1.28c). Hence, the condition (1.26b) can be guaranteed by the expected cost constraint

$$E\gamma\Big(y\big(a(\omega), x\big)\Big) \leq \gamma_0 \alpha^{\max}. \tag{1.29c}$$

Example 1.3 If the loss function $\gamma(y)$ is defined by a vector of individual loss functions γ_i for each state function $y_i = y_i(a, x)$, $i = 1, \ldots, m_y$, hence,

$$\gamma(y) = \Big(\gamma_1(y_1), \ldots, \gamma_{m_y}(y_{m_y})\Big)^T, \tag{1.30a}$$

then

$$\Gamma(x) = \left(\Gamma_1(x), \ldots, \Gamma_{m_y}(x)\right)^T, \ \Gamma_i(x) := E\gamma_i\left(y_i\left(a(\omega), x\right)\right), 1 \leq i \leq m_y, \tag{1.30b}$$

i.e., the m_y state functions $y_i, i = 1, \ldots, m_y$, will be treated separately.

Working with the more general expected failure or recourse cost functions $\Gamma = \Gamma(x)$, instead of (1.25a)–(1.25c), (1.26a)–(1.26c) and (1.27a)–(1.27c) we have the related substitute problems:

(I) *Expected total cost minimization*

$$\min \quad c_G E G_0\left(a(\omega), x\right) + c_f^T \Gamma(x), \tag{1.31a}$$

$$\text{s.t.} \quad x \in D. \tag{1.31b}$$

(II) *Expected primary cost minimization under expected failure or recourse cost constraints*

$$\min \quad E G_0\left(a(\omega), x\right) \tag{1.32a}$$

$$\text{s.t.}$$

$$\Gamma(x) \leq \Gamma^{\max} \tag{1.32b}$$

$$x \in D, \tag{1.32c}$$

(III) *Expected failure or recourse cost minimization under expected primary cost constraints*

$$\min \quad \Gamma(x) \tag{1.33a}$$

$$\text{s.t.}$$

$$E G_0\left(a(\omega), x\right) \leq G^{\max} \tag{1.33b}$$

$$x \in D. \tag{1.33c}$$

Here, c_G, c_f are (vectorial) weight coefficients, $\Gamma^{\max}$ is the vector of upper loss bounds, and "min" indicates again that $\Gamma(x)$ may be a vector valued function.

1.5 Basic Properties of Deterministic Substitute Problems

As can be seen from the conversion of an optimization problem with random parameters into a deterministic substitute problem, cf. Sect. 1.4.1, a central role is played by expectation or mean value functions of the type

$$\Gamma(x) = E\gamma\left(y\Big(a(\omega), x\Big)\right), \qquad x \in D_0, \tag{1.34a}$$

or more general

$$\Gamma(x) = Eg\Big(a(\omega), x\Big), \qquad x \in D_0. \tag{1.34b}$$

Here, $a = a(\omega)$ is a random ν-vector, $y = y(a, x)$ is an m_y-vector valued function on a certain subset of $\mathbb{R}^\nu \times \mathbb{R}^r$, and $\gamma = \gamma(z)$ is a real-valued function on a certain subset of $\mathbb{R}^{m_y}$.

Furthermore, $g = g(a, x)$ denotes a real-valued function on a certain subset of $\mathbb{R}^\nu \times \mathbb{R}^r$. In the following we suppose that the expectation in (1.34a)–(1.34b) exists and is finite for all input vectors x lying in an appropriate set $D_0 \subset \mathbb{R}^r$, cf. [7].

The following basic properties of the mean value functions Γ are needed in the following again and again.

Lemma 1.1 (Convexity) *Suppose that* $x \to g\Big(a(\omega), x\Big)$ *is convex a.s. (almost sure) on a fixed convex domain* $D_0 \subset \mathbb{R}^r$. *If* $Eg\Big(a(\omega), x\Big)$ *exists and is finite for each* $x \in D_0$, *then* $\Gamma = \Gamma(x)$ *is convex on* D_0.

Proof This property follows [20, 21, 27] directly from the linearity of the expectation operator. □

If $g = g(a, x)$ is defined by $g(a, x) := \gamma\Big(y(a, x)\Big)$, see (1.34a), then the above theorem yields the following result:

Corollary 1.1 *Suppose that* γ *is convex and* $E\gamma\left(y\Big(a(\omega), x\Big)\right)$ *exists and is finite for each* $x \in D_0$.

(a) If $x \to y\Big(a(\omega), x\Big)$ *is linear a.s., then* $\Gamma = \Gamma(x)$ *is convex.*

(b) If $x \to y\Big(a(\omega), x\Big)$ *is convex a.s., and* γ *is a convex, monotoneous nondecreasing function, then* $\Gamma = \Gamma(x)$ *is convex.*

It is well known [25] that a convex function is continuous on each open subset of its domain. A general sufficient condition for the continuity of Γ is given next.

Lemma 1.2 (Continuity) *Suppose that* $Eg\Big(a(\omega), x\Big)$ *exists and is finite for each* $x \in D_0$, *and assume that* $x \to g\Big(a(\omega), x\Big)$ *is continuous at* $x_0 \in D_0$ *a.s.. If there is a function* $\psi = \psi\Big(a(\omega)\Big)$ *having finite expectation such that*

$$\Big|g\Big(a(\omega), x\Big)\Big| \le \psi\Big(a(\omega)\Big) \text{ a.s. for all } x \in U(x_0) \cap D_0, \tag{1.35}$$

where $U(x_0)$ *is a neighborhood of* x_0, *then* $\Gamma = \Gamma(x)$ *is continuous at* x_0.

Proof The assertion can be shown by using Lebesgue's dominated convergence theorem, see, e.g., [27]. □

For the consideration of the differentiability of $\Gamma = \Gamma(x)$, let D denote an open subset of the domain D_0 of Γ.

Lemma 1.3 (Differentiability) *Suppose that*

(i) $Eg\Big(a(\omega), x\Big)$ *exists and is finite for each* $x \in D_0$,

(ii) $x \to g\Big(a(\omega), x\Big)$ *is differentiable on the open subset* D *of* D_0 *a.s. and*

(iii)

$$\left\| \nabla_x g\Big(a(\omega), x\Big) \right\| \le \psi\Big(a(\omega)\Big), \quad x \in D, \ a.s., \tag{1.36a}$$

where $\psi = \psi\Big(a(\omega)\Big)$ *is a function having finite expectation. Then the expectation of* $\nabla_x g\Big(a(\omega), x\Big)$ *exists and is finite,* $\Gamma = \Gamma(x)$ *is differentiable on* D *and*

$$\nabla\Gamma(x) = \nabla_x Eg\Big(a(\omega), x\Big) = E\nabla_x g\Big(a(\omega), x\Big), \quad x \in D. \tag{1.36b}$$

Proof Considering the difference quotients $\dfrac{\Delta\Gamma}{\Delta x_k}$, $k = 1, \ldots, r$, of Γ at a fixed point $x_0 \in D$, the assertion follows by means of the mean value theorem, inequality (1.36a) and Lebesgue's dominated convergence theorem, cf. [20, 21, 27]. □

Example 1.4 In case (1.34a), under obvious differentiability assumptions concerning γ and y we have $\nabla_x g(a, x) = \nabla_x y(a, x)^T \nabla\gamma\Big(y(a, x)\Big)$, where $\nabla_x y(a, x)$ denotes the Jacobian of $y = y(a, x)$ with respect to a. Hence, if (1.36b) holds, then

$$\nabla\Gamma(x) = E\nabla_x y\Big(a(\omega), x\Big)^T \nabla\gamma\left(y\Big(a(\omega), x\Big)\right). \tag{1.36c}$$

1.6 Approximations of Deterministic Substitute Problems in Optimal Design/Decision

The main problem in solving the deterministic substitute problems defined above is that the arising probability and expected cost functions $p_f = p_f(x), \Gamma = \Gamma(x), x \in \mathbb{R}^r$, are defined by means of multiple integrals over a ν-dimensional space.

Thus, the substitute problems may be solved, in practice, only by some approximative analytical and numerical methods [16, 20, 27, 33]. In the following we consider possible approximations for substitute problems based on general expected recourse cost functions $\Gamma = \Gamma(x)$ according to (1.34a) having a real-valued convex loss function $\gamma(z)$. Note that the probability of failure function $p_f = p_f(x)$ may be

approximated from above, see (1.29a)–(1.29b), by expected cost functions $\Gamma = \Gamma(x)$ having a nonnegative function $\gamma = \gamma(z)$ being bounded from below on the failure domain B^c. In the following several basic approximation methods are presented.

1.6.1 Approximation of the Loss Function

Suppose here that $\gamma = \gamma(y)$ is a continuously differentiable, convex loss function on $\mathbb{R}^{m_y}$. Let then denote

$$\overline{y}(x) := Ey\Big(a(\omega), x\Big) = \Big(Ey_1\Big(a(\omega), x\Big), \ldots, Ey_{m_y}\Big(a(\omega), x\Big)\Big)^T \tag{1.37}$$

the expectation of the vector $y = y\Big(a(\omega), x\Big)$ of state functions $y_i = y_i\Big(a(\omega), x\Big)$, $i = 1, \ldots, m_y$.

For an arbitrary continuously differentiable, convex loss function γ we have

$$\gamma\left(y\Big(a(\omega), x\Big)\right) \geq \gamma\Big(\overline{y}(x)\Big) + \nabla\gamma\Big(\overline{y}(x)\Big)^T\left(y\Big(a(\omega), x\Big) - \overline{y}(x)\right). \tag{1.38a}$$

Thus, taking expectations in (1.38a), we find Jensen's inequality

$$\Gamma(x) = E\gamma\left(y\Big(a(\omega), x\Big)\right) \geq \gamma\Big(\overline{y}(x)\Big) \tag{1.38b}$$

which holds for any convex function γ. Using the mean value theorem, we have

$$\gamma(y) = \gamma(\overline{y}) + \nabla\gamma(\hat{y})^T(y - \overline{y}), \tag{1.38c}$$

where $\hat{y}$ is a point on the line segment $\overline{y}y$ between $\overline{y}$ and y. By means of (1.38b), (1.38c) we get

$$0 \leq \Gamma(x) - \gamma\Big(\overline{y}(x)\Big) \leq E\left\|\nabla\gamma\left(\hat{y}\Big(a(\omega), x\Big)\right)\right\| \cdot \left\|y\Big(a(\omega), x\Big) - \overline{y}(x)\right\|. \tag{1.38d}$$

(a) Bounded gradient

If the gradient $\nabla\gamma$ is bounded on convex hull $R^{conv}(y(\cdot, \cdot))$ of the range of $y = y\Big(a(\omega), x\Big)$, $\omega \in \Omega, x \in D$, i.e., if

$$\|\nabla\gamma(y)\| \leq \vartheta^{\max} \quad \text{for each } y \in R^{conv}(y(\cdot, \cdot)), \tag{1.39a}$$

with a constant $\vartheta^{\max} > 0$, then

$$0 \le \Gamma(x) - \gamma\Big(\overline{y}(x)\Big) \le \vartheta^{\max} E \left\| y\Big(a(\omega), x\Big) - \overline{y}(x) \right\|, \quad x \in D. \tag{1.39b}$$

Since $t \to \sqrt{t}$, $t \ge 0$, is a concave function, we get

$$0 \le \Gamma(x) - \gamma\Big(\overline{y}(x)\Big) \le \vartheta^{\max} \sqrt{q(x)}\,, \tag{1.39c}$$

where

$$q(x) := E \left\| y\Big(a(\omega), x\Big) - \overline{y}(x) \right\|^2 = tr\, Q(x) \tag{1.39d}$$

is the generalized variance, and

$$Q(x) := \operatorname{cov}\Big(y\Big(a(\cdot), x\Big)\Big) \tag{1.39e}$$

denotes the covariance matrix of the random vector $y = y\Big(a(\omega), x\Big)$. Consequently, the expected loss function $\Gamma(x)$ can be approximated from above by

$$\Gamma(x) \le \gamma\Big(\overline{y}(x)\Big) + \vartheta^{\max} \sqrt{q(x)} \quad \text{for } x \in D. \tag{1.39f}$$

(b) Bounded eigenvalues of the Hessian

Considering second-order expansions of γ, with a vector $\tilde{y} \in \bar{y}y$ we find

$$\gamma(y) - \gamma(\overline{y}) = \nabla\gamma(\overline{y})^T (y - \overline{y}) + \frac{1}{2}(y - \overline{y})^T \nabla^2\gamma(\tilde{y})(y - \overline{y}). \tag{1.40a}$$

Suppose that the eigenvalues λ of $\nabla^2\gamma(y)$ are bounded from below and above on the convex hull $R^{conv}(y(\cdot,\cdot))$ of the range of $y = y\Big(a(\omega), x\Big)$ for all $\omega \in \Omega$, $x \in D$, i.e.,

$$0 < \lambda^{\min} \le \lambda\left(\nabla^2\gamma(y)\right) \le \lambda^{\max} < +\infty, \text{ for each } y \in R^{conv}(y(\cdot,\cdot)), \tag{1.40b}$$

with constants $0 < \lambda^{\min} \le \lambda^{\max}$. Taking expectations in (1.40a), we get

$$\gamma\Big(\overline{y}(x)\Big) + \frac{\lambda^{\min}}{2} q(x) \le \Gamma(x) \le \gamma\Big(\overline{y}(x)\Big) + \frac{\lambda^{\max}}{2} q(x), \quad x \in D. \tag{1.40c}$$

Consequently, using (1.39f) or (1.40c), various approximations for the deterministic substitute problems, (1.31a), (1.31b), (1.32a)–(1.32c), (1.33a)–(1.33c) may be obtained.

Based on the above approximations of expected cost functions, we state the following two approximates to (1.32a)–(1.32c), (1.33a)–(1.33c), resp., which are well known in *robust optimal design:*

(i) *Expected primary cost minimization under approximate expected failure or recourse cost constraints*

$$\min \quad EG_0\Big(a(\omega), x\Big) \tag{1.41a}$$

s.t.

$$\gamma\Big(\overline{y}(x)\Big) + c_0 q(x) \leq \Gamma^{\max} \tag{1.41b}$$

$$x \in D, \tag{1.41c}$$

where c_0 is a scale factor, cf. (1.39f) and (1.40c);

(ii) *Approximate expected failure or recourse cost minimization under expected primary cost constraints*

$$\min \quad \gamma\Big(\overline{y}(x)\Big) + c_0 q(x) \tag{1.42a}$$

s.t.

$$EG_0\Big(a(\omega), x\Big) \leq G^{\max} \tag{1.42b}$$

$$x \in D. \tag{1.42c}$$

Obviously, by means of (1.41a)–(1.41c) or (1.42a)–(1.42c) optimal designs x^* are achieved which

- yield a high mean performance of the structure/structural system
- are minimally sensitive or have a limited sensitivity with respect to random parameter variations (material, load, manufacturing, process, etc.) and
- cause only limited costs for design, construction, maintenance, etc.

1.6.2 Approximation of State (Performance) Functions

The numerical solution is simplified considerably if one can work with one single state function $y = y(a, x)$. Formally, this is possible by defining the function

$$y^{\min}(a, x) := \min_{1 \leq i \leq m_y} y_i(a, x). \tag{1.43a}$$

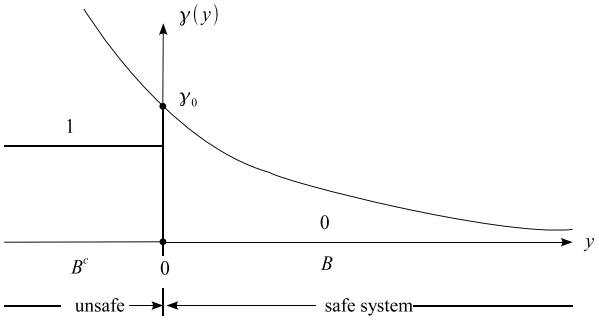


Fig. 1.1 Loss function γ

Indeed, according to (1.18b), (1.18c) the failure of the structure, the system can be represented by the condition

$$y^{\min}(a, x) \leq 0. \tag{1.43b}$$

Thus, the weakness or failure of the technical or economic device can be evaluated numerically by the function

$$\Gamma(x) := E\gamma\left(y^{\min}\Big(a(\omega), x\Big)\right) \tag{1.43c}$$

with a non-increasing loss function $\gamma : \mathbb{R} \to \mathbb{R}_+$, see Fig. 1.1.

However, the "min"-operator in (1.43a) yields a nonsmooth function $y^{\min} = y^{\min}(a, x)$ in general, and the straightforward computation of the mean and variance function

$$\overline{y^{\min}}(x) := Ey^{\min}\Big(a(\omega), x\Big) \tag{1.43d}$$

$$\sigma^2_{y^{\min}}(x) := \mathrm{Var}\Big(y^{\min}\Big(a(\cdot), x\Big)\Big) \tag{1.43e}$$

by means of Taylor expansion with respect to the model parameter vector a at $\overline{a} = Ea(\omega)$ is not possible, cf. Sect. 1.6.3.

According to the definition (1.43a), an upper bound for $\overline{y^{\min}}(x)$ is given by

$$\overline{y^{\min}}(x) \leq \min_{1\leq i\leq m_y} \overline{y}_i(x) = \min_{1\leq i\leq m_y} Ey_i\Big(a(\omega), x\Big).$$

Further approximations of $y^{\min}(a, x)$ and its moments can be found by using the representation

$$\min(a, b) = \frac{1}{2}\Big(a + b - |a - b|\Big)$$

of the minimum of two numbers $a, b \in \mathbb{R}$. For example, for an even index m_y we have

$$\begin{aligned} y^{\min}(a,x) &= \min_{i=1,3,\ldots,m_y-1} \min\Big(y_i(a,x),\, y_{i+1}(a,x)\Big) \\ &= \min_{i=1,3,\ldots,m_y-1} \frac{1}{2}\Big(y_i(a,x) + y_{i+1}(a,x) - \big|y_i(a,x) - y_{i+1}(a,x)\big|\Big). \end{aligned}$$

In many cases we may suppose that the state (performance) functions $y_i = y_i(a,x)$, $i = 1, \ldots, m_y$, are bounded from below, hence,

$$y_i(a,x) > -A, \quad i = 1, \ldots, m_y,$$

for all (a,x) under consideration with a positive constant $A > 0$. Thus, defining

$$\tilde{y}_i(a,x) := y_i(a,x) + A, \quad i = 1, \ldots, m_y,$$

and therefore

$$\tilde{y}^{\min}(a,x) := \min_{1 \le i \le m_y} \tilde{y}_i(a,x) = y^{\min}(a,x) + A,$$

we have

$$y^{\min}(a,x) \le 0 \quad \text{if and only if} \quad \tilde{y}^{\min}(a,x) \le A.$$

Hence, the survival/failure of the system or structure can also be studied by means of the positive function $\tilde{y}^{\min} = \tilde{y}^{\min}(a,x)$. Using now the theory of power or Hölder means [12], the minimum $\tilde{y}^{\min}(a,x)$ of positive functions can be represented also by the limit

$$\tilde{y}^{\min}(a,x) = \lim_{\lambda \to -\infty} \left(\frac{1}{m_y} \sum_{i=1}^{m_y} \tilde{y}_i(a,x)^{\lambda}\right)^{1/\lambda}$$

of the decreasing family of power means

$$M^{[\lambda]}(\tilde{y}) := \left(\frac{1}{m_y} \sum_{i=1}^{m_y} \tilde{y}_i^{\lambda}\right)^{1/\lambda}, \quad \lambda < 0.$$

Consequently, for each fixed $p > 0$ we also have

$$\tilde{y}^{\min}(a,x)^p = \lim_{\lambda \to -\infty} \left(\frac{1}{m_y} \sum_{i=1}^{m_y} \tilde{y}_i(a,x)^{\lambda}\right)^{p/\lambda}.$$

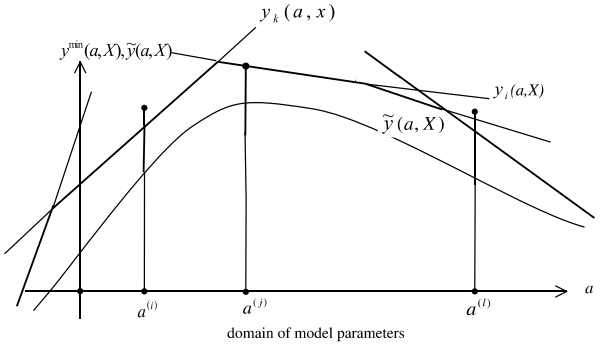


Fig. 1.2 Approximation $\tilde{y}(a,x)$ of $y^{\min}(a,x)$ for given x

Assuming that the expectation $EM^{[\lambda]}\Big(\tilde{y}\big(a(\omega)\big),x\Big)^p$ exists for an exponent $\lambda = \lambda_0 < 0$, by means of Lebesgue's bounded convergence theorem we get the moment representation

$$E\tilde{y}^{\min}\Big(a(\omega),x\Big)^p = \lim_{\lambda\to-\infty} E\left(\frac{1}{m_y}\sum_{i=1}^{m_y}\tilde{y}_i\Big(a(\omega),x\Big)^\lambda\right)^{p/\lambda}.$$

Since $t \to t^{p/\lambda}, t > 0$, is convex for each fixed $p > 0$ and $\lambda < 0$, by Jensen's inequality we have the lower moment bound

$$E\tilde{y}^{\min}\Big(a(\omega),x\Big)^p \geq \lim_{\lambda\to-\infty}\left(\frac{1}{m_y}\sum_{i=1}^{m_y}E\tilde{y}_i\Big(a(\omega),x\Big)^\lambda\right)^{p/\lambda}.$$

Hence, for the pth order moment of $\tilde{y}^{\min}\Big(a(\cdot),x\Big)$ we get the approximations

$$E\left(\frac{1}{m_y}\sum_{i=1}^{m_y}\tilde{y}_i\Big(a(\omega),x\Big)\right)^{p/\lambda} \geq \left(\frac{1}{m_y}\sum_{i=1}^{m_y}E\tilde{y}_i\Big(a(\omega),x\Big)^\lambda\right)^{p/\lambda}$$

for some $\lambda < 0$.

Using regression techniques, Response Surface Methods (RSM), etc., for given vector x, the function $a \to y^{\min}(a,x)$ can be approximated [8, 11, 19, 24, 42] by functions $\widetilde{y} = \widetilde{y}(a,x)$ being sufficiently smooth with respect to the parameter vector a (Fig. 1.2).

In many important cases, for each $i = 1,\ldots,m_y$, the state functions

$$(a,x) \longrightarrow y_i(a,x)$$

are bilinear functions. Thus, in this case $y^{\min} = y^{\min}(a, x)$ is a piecewise linear function with respect to a. Fitting a linear or quadratic Response Surface Model [9, 10, 35, 36]

$$\widetilde{y}(a, x) := c(x) + q(x)^T(a - \overline{a}) + (a - \overline{a})^T Y(x)(a - \overline{a}) \tag{1.43f}$$

to $a \to y^{\min}(a, x)$, after the selection of appropriate reference points

$$a^{(j)} := \overline{a} + d_a^{(j)},\ j = 1, \ldots, p, \tag{1.43g}$$

with "design" points $d_a^{(j)} \in \mathbb{R}^\nu$, $j = 1, \ldots, p$, the unknown coefficients $c = c(x)$, $q = q(x)$ and $Y = Y(x)$ are obtained by minimizing the mean square error

$$\rho(c, q, Y) := \sum_{j=1}^{p} \left(\widetilde{y}(a^{(j)}, x) - y^{\min}(a^{(j)}, x)\right)^2 \tag{1.43h}$$

with respect to (c, q, Y). Since the model (1.43f) depends linearly on the function parameters (c, q, Y), explicit formulas for the optimal coefficients

$$c^* = c^*(x), q^* = q^*(x), Y^* = Y^*(x) \tag{1.43i}$$

are obtained from this least squares estimation method, cf. [33].

1.6.2.1 Approximation of Expected Loss Functions

Corresponding to the approximation (1.43f) of $y^{\min} = y^{\min}(a, x)$, using again least squares techniques, a mean value function $\Gamma(x) = E\gamma\left(y\Big(a(\omega), x\Big)\right)$, cf. (1.28a), can be approximated at a given point $x_0 \in \mathbb{R}^\nu$ by a linear or quadratic Response Surface Function

$$\widetilde{\Gamma}(x) := \beta_0 + \beta_I^T(x - x_0) + (x - x_0)^T B(x - x_0), \tag{1.43j}$$

with scalar, vector and matrix parameters β_0, β_I, B. In this case estimates $y^{(i)} = \hat{\Gamma}^{(i)}$ of $\Gamma(x)$ are needed at some reference points $x^{(i)} = x_0 + d^{(i)}$, $i = 1, \ldots, p$. Details are given in [33].

1.6.3 Taylor Expansion Methods

As can be seen above, cf. (1.34a)–(1.34b), in the objective and/or in the constraints of substitute problems for optimization problems with random data mean value functions of the type

$$\Gamma(x) := Eg\Big(a(\omega), x\Big)$$

occur. Here, $g = g(a, x)$ is a real-valued function on a subset of $\mathbb{R}^\nu \times \mathbb{R}^r$, and $a = a(\omega)$ is a random ν vector.

1.6.3.1 (Complete) Expansion with Respect to *a*

Suppose that on its domain the function $g = g(a, x)$ has partial derivatives $\nabla_a^l g(a, x), l = 0, 1, \ldots, l_g + 1$, up to order $l_g + 1$. Note that the gradient $\nabla_a g(a, x)$ contains the so-called *sensitivities* $\dfrac{\partial g}{\partial a_j}(a, x)$, $j = 1, \ldots, \nu$, of g with respect to the parameter vector a at (a, x). In the same way, the higher order partial derivatives $\nabla_a^l g(a, x), l > 1$, represent the *higher order sensitivities* of g with respect to a at (a, x). Taylor expansion of $g = g(a, x)$ with respect to a at $\overline{a} := Ea(\omega)$ yields

$$g(a, x) = \sum_{l=0}^{l_g} \frac{1}{l!} \nabla_a^l g(\bar{a}, x) \cdot (a - \bar{a})^l + \frac{1}{(l_g + 1)!} \nabla_a^{l_g+1} g(\hat{a}, x) \cdot (a - \bar{a})^{l_g+1}, \tag{1.44a}$$

where $\hat{a} := \bar{a} + \vartheta(a - \bar{a}), 0 < \vartheta < 1$, and $(a - \bar{a})^l$ denotes the system of l-th order products

$$\prod_{j=1}^{\nu} (a_j - \bar{a}_j)^{l_j}$$

with $l_j \in \mathbb{N} \cup \{0\}$, $j = 1, \ldots, \nu, l_1 + l_2 + \ldots + l_\nu = l$. If $g = g(a, x)$ is defined by

$$g(a, x) := \gamma\Big(y(a, x)\Big),$$

see (1.34a), then the partial derivatives $\nabla_a^l g$ of g up to the second-order read

$$\nabla_a g(a, x) = \Big(\nabla_a y(a, x)\Big)^T \nabla\gamma\Big(y(a, x)\Big) \tag{1.44b}$$

$$\nabla_a^2 g(a, x) = \Big(\nabla_a y(a, x)\Big)^T \nabla^2\gamma\Big(y(a, x)\Big)\nabla_a y(a, x) + \nabla\gamma\Big(y(a, x)\Big) \cdot \nabla_a^2 y(a, x), \tag{1.44c}$$

where

$$(\nabla\gamma)\cdot\nabla_a^2 y := \left((\nabla\gamma)^T \frac{\partial^2 y}{\partial a_k \partial a_l}\right)_{k,l=1,\ldots,v}. \tag{1.44d}$$

Taking expectations in (1.44a), $\Gamma(x)$ can be approximated, cf. Sect. 1.6.1, by

$$\widetilde{\Gamma}(x) := g(\bar{a}, x) + \sum_{l=2}^{l_g} \nabla_a^l g(\bar{a}, x)\cdot E\Big(a(\omega)-\bar{a}\Big)^l, \tag{1.45a}$$

where $E\Big(a(\omega)-\bar{a}\Big)^l$ denotes the system of mixed lth central moments of the random vector $a(\omega) = \Big(a_1(\omega), \ldots, a_v(\omega)\Big)^T$. Assuming that the domain of $g = g(a, x)$ is convex with respect to a, we get the error estimate

$$\Big|\Gamma(x) - \widetilde{\Gamma}(x)\Big| \le \frac{1}{(l_g+1)!} E \sup_{0\le\vartheta\le 1} \Big\|\nabla_a^{l_g+1} g\Big(\bar{a}+\vartheta\big(a(\omega)-\bar{a}\big), x\Big)\Big\| \times \Big\|a(\omega)-\bar{a}\Big\|^{l_g+1}. \tag{1.45b}$$

In many practical cases the random parameter v-vector $a = a(\omega)$ has a convex, bounded support, and $\nabla_a^{l_g+1} g$ is continuous. Then the L_∞-norm

$$r(x) := \frac{1}{(l_g+1)!} \operatorname*{ess\,sup}_{\omega\in\Omega} \Big\|\nabla_a^{l_g+1} g\Big(a(\omega), x\Big)\Big\| \tag{1.45c}$$

is finite for all x under consideration, and (1.45b), (1.45c) yield the error bound

$$\Big|\Gamma(x) - \widetilde{\Gamma}(x)\Big| \le r(x) E\Big\|a(\omega)-\bar{a}\Big\|^{l_g+1}. \tag{1.45d}$$

Remark 1.4 The above-described method can be extended to the case of vector valued loss functions $\gamma(z) = \Big(\gamma_1(z), \ldots, \gamma_{m_\gamma}(z)\Big)^T$.

1.6.3.2 Inner (Partial) Expansions with Respect to *a*

In generalization of (1.34a), in many cases $\Gamma(x)$ is defined by

$$\Gamma(x) = E\gamma\Big(a(\omega), y\Big(a(\omega), x\Big)\Big), \tag{1.46a}$$

hence, the loss function $\gamma = \gamma(a, y)$ depends also explicitly on the parameter vector a. This may occur, e.g., in case of randomly varying cost factors.

Linearizing now the vector function $y = y(a, x)$ with respect to a at $\bar{a}$, thus,

$$y(a, x) \approx y_{(1)}(a, x) := y(\bar{a}, x) + \nabla_a y(\bar{a}, x)(a - \bar{a}), \tag{1.46b}$$

the mean value function $\Gamma(x)$ is approximated by

$$\widetilde{\Gamma}(x) := E\gamma\left(a(\omega), y(\bar{a}, x) + \nabla_a y(\bar{a}, x)\Big(a(\omega) - \bar{a}\Big)\right). \tag{1.46c}$$

This approximation is very advantageous in case that the cost function $\gamma = \gamma(a, y)$ is a *quadratic function in* y. In case of a cost function $\gamma = \gamma(a, y)$ being linear in the vector y, also *quadratic expansions of* $y = y(a, x)$ *with respect to* a many be taken into account.

Corresponding to (1.37), (1.39e), define

$$\overline{y}_{(1)}(x) := Ey_{(1)}\Big(a(\omega), x\Big) = y(\overline{a}, x) \tag{1.46d}$$

$$Q_{(1)}(x) := \operatorname{cov}\left(y_{(1)}\Big(a(\cdot), x\Big)\right) = \nabla_a y(\overline{a}, x)\operatorname{cov}\Big(a(\cdot)\Big)\nabla_a y(\overline{a}, x)^T. \tag{1.46e}$$

In case of convex loss functions γ, approximates of $\widetilde{\Gamma}$ and the corresponding substitute problems based on $\widetilde{\Gamma}$ may be obtained now by applying the methods described in Sect. 1.6.1 Explicit representations for $\widetilde{\Gamma}$ are obtained in case of quadratic loss functions γ.

Error estimates can be derived easily for Lipschitz(L)-continuous or convex loss function γ. In case of a Lipschitz-continuous loss function $\gamma(a, \cdot)$ with Lipschitz constant $L = L(a) > 0$, e.g., for sublinear [27, 28] loss functions, using (1.46d) we have

$$\left|\Gamma(x) - \widetilde{\Gamma}(x)\right| \le L_0 \cdot E\left\|y\Big(a(\omega), x\Big) - y_{(1)}\Big(a(\omega), x\Big)\right\|, \tag{1.46f}$$

provided that L_0 denotes a finite upper bound of the L-constants $L = L(a)$.

Applying the mean value theorem [15], under appropriate second-order differentiability assumptions, for the right-hand side of (1.46f) we find the following stochastic version of the mean value theorem

$$\begin{aligned} &E\left\|y\Big(a(\omega), x\Big) - y_{(1)}\Big(a(\omega), x\Big)\right\| \\ &\le E\left\|a(\omega) - \overline{a}\right\|^2 \sup_{0\le\vartheta\le 1}\left\|\nabla_a^2 y\left(\overline{a} + \vartheta\Big(a(\omega) - \overline{a}\Big), x\right)\right\|. \end{aligned} \tag{1.46g}$$

1.7 Approximation of Probabilities—Probability Inequalities

In reliability analysis of engineering/economic structures or systems, a main problem is the computation of probabilities

$$P\left(\bigcup_{i=1}^{N} V_i\right) := P\left(a(\omega) \in \bigcup_{i=1}^{N} V_i\right) \tag{1.47a}$$

or

$$P\left(\bigcap_{j=1}^{N} S_j\right) := P\left(a(\omega) \in \bigcap_{j=1}^{N} S_j\right) \tag{1.47b}$$

of unions and intersections of certain failure/survival domains (events) V_j, S_j, $j = 1, \ldots, N$. These domains (events) arise from the representation of the structure or system by a combination of certain series and/or parallel substructures/systems. Due to the high complexity of the basic physical relations, several approximation techniques are needed for the evaluation of (1.47a), (1.47b).

1.7.1 Bonferroni-Type Inequalities

In the following $V_1, V_2, \ldots, V_N$ denote arbitrary (Borel-)measurable subsets of the parameter space $\mathbb{R}^\nu$, and the abbreviation

$$P(V) := P\Big(a(\omega) \in V\Big) \tag{1.47c}$$

is used for any measurable subset V of $\mathbb{R}^\nu$.

Starting from the representation of the probability of a union of N events,

$$P\left(\bigcup_{j=1}^{N} V_j\right) = \sum_{k=1}^{N} (-1)^{k-1} s_{k,N}, \tag{1.48a}$$

where

$$s_{k,N} := \sum_{1 \le i_1 < i_2 < \cdots < i_k \le N} P\left(\bigcap_{l=1}^{k} V_{i_l}\right), \tag{1.48b}$$

we obtain [17] the well-known basic Bonferroni bounds

$$P\left(\bigcup_{j=1}^{N} V_j\right) \leq \sum_{k=1}^{\rho} (-1)^{k-1} s_{k,N} \text{ for } \rho \geq 1, \rho \text{ odd} \tag{1.48c}$$

$$P\left(\bigcup_{j=1}^{N} V_j\right) \geq \sum_{k=1}^{\rho} (-1)^{k-1} s_{k,N} \text{ for } \rho \geq 1, \rho \text{ even}. \tag{1.48d}$$

Besides (1.48c), (1.48d), a large amount of related bounds of different complexity are available, cf. [17, 50]. Important bounds of first and second degree are given below:

$$\max_{1 \leq j \leq N} q_j \leq P\left(\bigcup_{j=1}^{N} V_j\right) \leq Q_1 \tag{1.49a}$$

$$Q_1 - Q_2 \leq P\left(\bigcup_{j=1}^{N} V_j\right) \leq Q_1 - \max_{1 \leq l \leq N} \sum_{i \neq l} q_{il} \tag{1.49b}$$

$$\frac{Q_1^2}{Q_1 + 2Q_2} \leq P\left(\bigcup_{j=1}^{N} V_j\right) \leq Q_1. \tag{1.49c}$$

The above quantities q_j, q_{ij}, Q_1, Q_2 are defined as follows:

$$Q_1 := \sum_{j=1}^{N} q_j \text{ with } q_j := P(V_j) \tag{1.49d}$$

$$Q_2 := \sum_{j=2}^{N} \sum_{i=1}^{j-1} q_{ij} \text{ with } q_{ij} := P(V_i \cap V_j). \tag{1.49e}$$

Moreover, defining

$$q := (q_1, \ldots, q_N),\ Q := (q_{ij})_{1 \leq i,j \leq N}, \tag{1.49f}$$

we have

$$P\left(\bigcup_{j=1}^{N} V_j\right) \geq q^T Q^- q, \tag{1.49g}$$

where Q^- denotes the generalized inverse of Q, cf. [50].

1.7.2 Tschebyscheff-Type Inequalities

In many cases the survival or feasible domain (event) $S = \bigcap_{i=1}^{m} S_i$ is represented by a certain number m of inequality constraints of the type

$$y_{li} < (\leq) y_i(a, x) < (\leq) y_{ui}, \quad i = 1, \ldots, m, \tag{1.50a}$$

as, e.g., operating conditions, behavioral constraints. Hence, for a fixed input, design or control vector x, the event $S = S(x)$ is given by

$$S := \{a \in \mathbb{R}^{\nu} : y_{li} < (\leq) y_i(a, x) < (\leq) y_{ui},\ i = 1, \ldots, m\}. \tag{1.50b}$$

Here,

$$y_i = y_i(a, x), \quad i = 1, \ldots, m \tag{1.50c}$$

are certain functions, e.g., response, output, or performance functions of the structure, system, defined on (a subset of) $\mathbb{R}^{\nu} \times \mathbb{R}^{r}$.

Moreover, $y_{li} < y_{ui}, i = 1, \ldots, m$, are lower and upper bounds for the variables $y_i, i = 1, \ldots, m$. In the case of one-sided constraints some bounds y_{li}, y_{ui} are infinite.

1.7.2.1 Two-Sided Constraints

If $y_{li} < y_{ui}, i = 1, \ldots, m$, are finite bounds, (1.50a) can be represented by

$$|y_i(a, x) - y_{ic}| < (\leq) \rho_i, \quad i = 1, \ldots, m, \tag{1.50d}$$

where the quantities $y_{ic}, \rho_i, i = 1, \ldots, m$, are defined by

$$y_{ic} := \frac{y_{li} + y_{ui}}{2}, \quad \rho_i := \frac{y_{ui} - y_{li}}{2}. \tag{1.50e}$$

Consequently, for the probability $P(S)$ of the event S, defined by (1.50b), we have

$$P(S) = P\left(\left|y_i\Big(a(\omega), x\Big) - y_{ic}\right| < (\leq) \rho_i,\ i = 1, \ldots, m\right). \tag{1.50f}$$

Introducing the random variables

$$\tilde{y}_i\Big(a(\omega), x\Big) := \frac{y_i\Big(a(\omega), x\Big) - y_{ic}}{\rho_i}, \quad i = 1, \ldots, m, \tag{1.51a}$$

and the set

$$B := \left\{y \in \mathbb{R}^m : |y_i| < (\leq) 1,\ i = 1, \ldots, m\right\}, \tag{1.51b}$$

with $\tilde{y} = (\tilde{y}_i)_{1 \le i \le m}$, we get

$$P(S) = P\left(\tilde{y}\Big(a(\omega), x\Big) \in B\right). \tag{1.51c}$$

Considering any (measurable) function $\varphi : \mathbb{R}^m \to \mathbb{R}$ such that

$$i)\varphi(y) \ge 0, \qquad y \in \mathbb{R}^m \tag{1.51d}$$

$$ii)\varphi(y) \ge \varphi_0 > 0, \qquad \text{if } y \notin B, \tag{1.51e}$$

with a positive constant φ_0, we find the following result:

Theorem 1.1 *For any (measurable) function $\varphi : \mathbb{R}^m \to \mathbb{R}$ fulfilling conditions (1.51d), (1.51e), the following* **Tschebyscheff-type inequality** *holds*

$$P\left(y_{li} < (\le) y_i\Big(a(\omega), x\Big) < (\le) y_{ui}, i = 1, \ldots, m\right) \ge 1 - \frac{1}{\varphi_0} E\varphi\left(\tilde{y}\Big(a(\omega), x\Big)\right), \tag{1.52}$$

provided that the expectation in (1.52) exists and is finite.

Proof If $P_{\tilde{y}\left(a(\cdot),x\right)}$ denotes the probability distribution of the random m-vector $\tilde{y} = \tilde{y}\Big(a(\omega), x\Big)$, then

$$\begin{aligned} E\varphi\left(\tilde{y}\Big(a(\omega), x\Big)\right) &= \int_{y \in B} \varphi(y) P_{\tilde{y}(a(\cdot),x)}(dy) + \int_{y \in B^c} \varphi(y) P_{\tilde{y}(a(\cdot),x)}(dy) \\ &\ge \int_{y \in B^c} \varphi(y) P_{\tilde{y}(a(\cdot),x)}(dy) \ge \varphi_0 \int_{y \in B^c} P_{\tilde{y}(a(\cdot),x)}(dy) \\ &= \varphi_0 P\left(\tilde{y}\Big(a(\omega), x\Big) \notin B\right) = \varphi_0\left(1 - P\left(\tilde{y}\Big(a(\omega), x\Big) \in B\right)\right), \end{aligned}$$

which yields the assertion, cf. (1.51c). □

Remark 1.5 Note that $P(S) \ge \alpha_s$ with a given minimum reliability $\alpha_s \in (0, 1]$ can be guaranteed by the expected cost constraint

$$E\varphi\left(\tilde{y}\Big(a(\omega), x\Big)\right) \le (1 - \alpha_s)\varphi_0.$$

Example 1.5 If $\varphi = 1_{B^c}$ is the indicator function of the complement B^c of B, then $\varphi_0 = 1$ and (1.52) holds with the equality sign (Fig. 1.3).

Example 1.6 For a given positive definite $m \times m$ matrix C, define $\varphi(y) := y^T C y$, $y \in \mathbb{R}^m$. Then, cf. (1.51b), (1.51d), (1.51e),

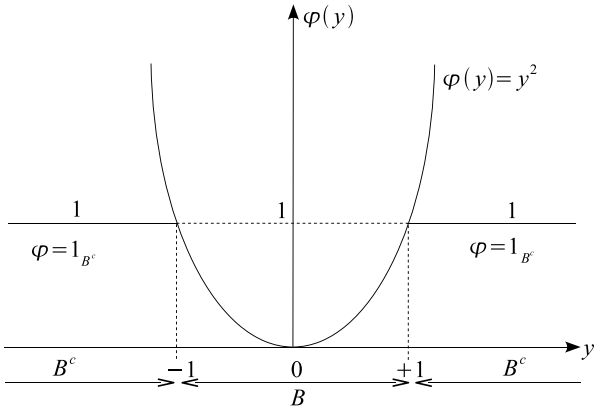


Fig. 1.3 Function $\varphi = \varphi(y)$

$$\min_{y \notin B} \varphi(y) = \min_{1 \le i \le m} \left\{ \min_{y_i \ge 1} y^T C y, \min_{y_i \le -1} y^T C y \right\}. \tag{1.53a}$$

Thus, the lower bound φ_0 follows by considering the convex optimization problems arising in the right-hand side of (1.53a). Moreover, the expectation $E\varphi(\tilde{y})$ needed in (1.52) is given, see (1.51a), by

$$\begin{aligned} &E\varphi(\tilde{y}) = E\tilde{y}^T C\tilde{y} = E\mathrm{tr} C\tilde{y}\tilde{y}^T, \\ &= \mathrm{tr} C(\mathrm{diag}\ \rho)^{-1} \left(\mathrm{cov}\ y\Big(a(\cdot), x\Big) + \Big(\overline{y}(x) - y_c\Big)\Big(\overline{y}(x) - y_c\Big)^T \right) (\mathrm{diag}\ \rho)^{-1}, \end{aligned} \tag{1.53b}$$

where "tr" denotes the trace of a matrix, diag ρ is the diagonal matrix diag $\rho := (\rho_i \delta_{ij})$, $y_c := (y_{ic})$, see (1.50e), and $\overline{y} = \overline{y}(x) := \Big(Ey_i\Big(a(\omega), x\Big)\Big)$. Moreover, for $E\varphi(\tilde{y})$ we have the upper bound

$$E\varphi(\tilde{y}) \le \|C\|\ \|(\mathrm{diag}\ \rho)^{-1}\|^2 \Big(tr\ \mathrm{cov} y\Big(a(\cdot), x\Big) + \|\overline{y}(x) - y_c\|^2 \Big). \tag{1.53c}$$

Example 1.7 Assuming in Example 1.6 that $C = \mathrm{diag}\ (c_{ii})$ is a diagonal matrix with positive elements $c_{ii} > 0, i = 1, \ldots, m$, then

$$\min_{y \notin B} \varphi(y) = \min_{1 \le i \le m} c_{ii} > 0, \tag{1.53d}$$

and $E\varphi(\tilde{y})$ is given by

$$E\varphi(\tilde{y}) = \sum_{i=1}^{m} c_{ii} \frac{E\left(y_i\left(a(\omega), x\right) - y_{ic}\right)^2}{\rho_i^2} \tag{1.53e}$$
$$= \sum_{i=1}^{m} c_{ii} \frac{\sigma_{y_i}^2\left(a(\cdot), x\right) + \left(\overline{y}_i(x) - y_{ic}\right)^2}{\rho_i^2}.$$

1.7.2.2 One-Sided Inequalities

Suppose that exactly one of the two bounds $y_{li} < y_{ui}$ is infinite for each $i = 1, \ldots, m$. Multiplying the corresponding constraints in (1.50a) by -1, the admissible domain $S = S(x)$, cf. (1.50b), can be represented by

$$S(x) = \{a \in \mathbb{R}^v : \tilde{y}_i(a, x) < (\leq)\ 0,\ i = 1, \ldots, m\}, \tag{1.54a}$$

where $\tilde{y}_i := y_i - y_{ui}$, if $y_{li} = -\infty$, and $\tilde{y}_i := y_{li} - y_i$, if $y_{ui} = +\infty$. If we set $\tilde{y}(a, x) := \left(\tilde{y}_i(a, x)\right)$ and

$$\tilde{B} := \left\{y \in \mathbb{R}^m : y_i < (\leq)\ 0,\ i = 1, \ldots, m\right\}, \tag{1.54b}$$

then

$$P\left(S(x)\right) = P\left(\tilde{y}\left(a(\omega), x\right) \in \tilde{B}\right). \tag{1.54c}$$

In this case, cf. (1.51d), (1.51e), also consider a function $\varphi : \mathbb{R}^m \to \mathbb{R}$ such that

$$i)\varphi(y) \geq 0, \qquad y \in \mathbb{R}^m \tag{1.55a}$$
$$ii)\varphi(y) \geq \varphi_0 > 0, \qquad \text{if } y \notin \tilde{B}. \tag{1.55b}$$

Then, corresponding to Theorem 1.1, we have this result:

Theorem 1.2 (Markov-type inequality) *If $\varphi : \mathbb{R}^m \to \mathbb{R}$ is any (measurable) function fulfilling conditions (1.55a,b), then*

$$P\left(\tilde{y}\left(a(\omega), x\right) < (\leq)\ 0\right) \geq 1 - \frac{1}{\varphi_0} E\varphi\left(\tilde{y}\left(a(\omega), x\right)\right), \tag{1.56}$$

provided that the expectation in (1.56) exists and is finite.

Remark 1.6 Note that a related inequality was already used in Example 1.2.

Example 1.8 If $\varphi(y) := \sum_{i=1}^{m} w_i e^{\alpha_i y_i}$ with positive constants $w_i, \alpha_i,\ i = 1, \ldots, m$, then

$$\inf_{y \notin \tilde{B}} \varphi(y) = \min_{1 \le i \le m} w_i > 0 \tag{1.57a}$$

and

$$E\varphi\left(\tilde{y}\left(a(\omega), x\right)\right) = \sum_{i=1}^{m} w_i E e^{\alpha_i \tilde{y}_i\left(a(\omega), x\right)}, \tag{1.57b}$$

where the expectation in (1.57b) can be computed approximately by Taylor expansion:

$$\begin{aligned} E e^{\alpha_i \tilde{y}_i} &= e^{\alpha_i \overline{\tilde{y}}_i(x)} E e^{\alpha_i \left(\tilde{y}_i - \overline{\tilde{y}}_i(x)\right)} \\ &\approx e^{\alpha_i \overline{\tilde{y}}_i(x)} \left(1 + \frac{\alpha_i^2}{2} E\left(y_i\left(a(\omega), x\right) - \overline{y}_i(x)\right)^2\right) \\ &= e^{\alpha_i \overline{\tilde{y}}_i(x)} \left(1 + \frac{\alpha_i^2}{2} \sigma^2_{y_i(a(\cdot), x)}\right). \end{aligned} \tag{1.57c}$$

Supposing that $y_i = y_i\left(a(\omega), x\right)$ is a normal distributed random variable, then

$$E e^{\alpha_i \tilde{y}_i} = e^{\alpha_i \overline{\tilde{y}}_i(x)} e^{\frac{1}{2}\alpha_i^2 \sigma^2_{y_i(a(\cdot), x)}}. \tag{1.57d}$$

Example 1.9 Consider $\varphi(y) := (y - b)^T C (y - b)$, where, cf. Example 1.6, C is a positive definite $m \times m$ matrix and $b < 0$ a fixed m-vector. In this case we again have

$$\min_{y \notin \tilde{B}} \varphi(y) = \min_{1 \le i \le m} \min_{y_i \ge 0} \varphi(y)$$

and

$$\begin{aligned} E\varphi(\tilde{y}) &= E\left(\tilde{y}\left(a(\omega), x\right) - b\right)^T C \left(\tilde{y}\left(a(\omega), x\right) - b\right) \\ &= \mathrm{tr} C E \left(\tilde{y}\left(a(\omega), x\right) - b\right) \left(\tilde{y}\left(a(\omega), x\right) - b\right)^T. \end{aligned}$$

Note that

$$\tilde{y}_i(a, x) - b_i = \begin{cases} y_i(a, x) - (y_{ui} + b_i), & \text{if } y_{li} = -\infty \\ y_{li} - b_i - y_i(a, x), & \text{if } y_{ui} = +\infty, \end{cases}$$

where $y_{ui} + b_i < y_{ui}$ and $y_{li} < y_{li} - b_i$.

Remark 1.7 The one-sided case can also be reduced approximatively to the two-sided case by selecting a sufficiently large, but finite upper bound $\tilde{y}_{ui} \in \mathbb{R}$, lower bound $\tilde{y}_{li} \in \mathbb{R}$, resp., if $y_{ui} = +\infty$, $y_{li} = -\infty$.

References

1. Augusti, G., et al.: Probabilistic Methods in Structural Engineering. Chapman and Hall, London (1984). https://doi.org/10.1201/9781482267457
2. Aurnhammer, A.: Optimale Stochastische Trajektorienplanung und Regelung von Industrierobotern. No. 1032 in Fortschrittberichte VDI, Reihe 8. VDI-Verlag GmbH, Düsseldorf (2004)
3. Bastian, G., et al.: Theory of Robot Control. Springer, Berlin (1996)
4. Bauer, H.: Wahrscheinlichkeitstheorie und Grundzüge der Masstheorie. Walter de Gruyter & Co., Berlin (1968)
5. Bauer, H.: Probability Theory. De Gruyter, Berlin (1996)
6. Berger, J.: Statistical Decision Theory and Bayesian Analysis. Springer, New York (1985)
7. Bertsekas, D.: Stochastic optimization problems with nondifferentiable cost functionals. J. Optim. Theory Appl. **12**(2), 218–231 (1973). https://doi.org/10.1007/BF00934819
8. Biles, W.E., Swain, J.J.: Mathematical programming and the optimization of computer simulations. In: R.S. Avriel M.and Dembo (ed.) Engineering Optimization, Mathematical Programming Studies, vol. 11, pp. 189–207. Springer, Berlin (1979)
9. Box, G., Draper, N.: Empirical Model-building and Response Surfaces. Wiley, New York (1987)
10. Box, G.E.P., Wilson, K.B.: On the experimental attainment of optimum conditions. J. Roy. Stat. Soc.: Ser. B (Methodol.) **13**(1), 1–38 (1951). https://doi.org/10.1111/j.2517-6161.1951.tb00067.x
11. Bucher, C., Bourgund, U.: A fast and efficient response surface approach for structural reliability problems. Struct. Saf. **7**(1), 57–66 (1990). https://doi.org/10.1016/0167-4730(90)90012-E
12. Bullen, P.: Handbook of Means and Their Inequalities. Kluwer Academic Publishing, Dordrecht (2003)
13. Chankong, V., Haimes, Y.Y.: Multiobjective Decision Making. North Holland, New York (1983)
14. Craig, J.: Adaptive Control of Mechanical Manipulators. Addison-Wesley (1988)
15. Dieudonné, J.: Foundations of Modern Analysis. Academic, New York (1969)
16. Ermoliev, Y., Wets, R. (eds.): Numerical Techniques for Stochastic Optimization, Springer Series in Computational Mathematics, vol. 10. Springer, Berlin (1988)
17. Galambos, J., Simonelli, I.: Bonferroni-Type Inequalities with Applications. Springer, New York (1996)
18. Gänssler, P., Stute, W.: Wahrscheinlichkeitstheorie. Springer, Berlin (1977)
19. Jacobson, S., Schruben, L.: Techniques for simulation response optimization. Oper. Res. Lett. **8**(1), 1–9 (1989). https://doi.org/10.1016/0167-6377(89)90025-4
20. Kall, P.: Stochastic Linear Programming. Springer, Berlin (1976)
21. Kall, P., Wallace, S.: Stochastic Programming. Stochastic Programming. Wiley, Chichester (1994)
22. Kesten, O., et al. (eds.): Review of Economic Design. Springer, Heidelberg
23. Kirsch, U.: Structural Optimization. Springer, Berlin (1993)
24. Kleijnen, J.: Statistical Tools for Simulation Practitioners. No. 76 in Statistics. Marcel Dekker (1987)
25. Luenberger, D.: Optimization by Vector Space Methods. Wiley, New York (1969)
26. Luenberger, D.: Introduction to Linear and Nonlinear Programming. Addison-Wesley Publishing Company, Reading (1973)
27. Marti, K.: Approximationen stochastischer Optimierungsprobleme. Hain Konigstein/Ts (1979)
28. Marti, K.: Optimierungsverfahren. Vorlesung an der Hochschule der Bundeswehr München (1983)
29. Marti, K.: Stochastic optimization methods in structural mechanics. ZAMM **70**, T 742–T745 (1990))
30. Marti, K.: Path planning for robots under stochastic uncertainty. Optimization **45**(1–4), 163–195 (1999). https://doi.org/10.1080/02331939908844432
31. Marti, K.: Stochastic optimization methods in optimal engineering design under stochastic uncertainty. ZAMM **83**(11), 1–18 (2003)

32. Marti, K.: Reliability analysis of technical systems/structures by means of polyhedral approximation of the safe/unsafe domain. GAMM Mitteilungen **30**(2), 211–254 (2007)
33. Marti, K.: Stochastic Optimization Methods, 2nd edn. Springer, Berlin (2008). https://doi.org/10.1007/978-3-540-79458-5
34. Marti, K., Ermoliev, Y., Pflug, G. (eds.): Dynamic Stochastic Optimization, LNEMS, vol. 532. Springer, Berlin (2004)
35. Montgomery, C.: Design and Analysis of Experiments. Wiley, New York (1984)
36. Myers, R.: Response Surface Methodology. Allyn and Bacon, Boston (1971)
37. Pfeiffer, F., Johanni, R.: A concept for manipulator trajectory planning. IEEE J. Robot. Autom. **3**(2), 115–123 (1987)
38. Prékopa, A., Szántai, T.: Flood control reservoir system design using stochastic programming. In: Balinski, M.L., Lemarechal, C. (eds.) Mathematical Programming in Use, pp. 138–151. Springer, Berlin (1978)
39. Rackwitz, R., Cuntze, R.: Formulations of reliability-oriented optimization. Eng. Optim. **11**(1–2), 69–76 (1987). https://doi.org/10.1080/03052158708941037
40. Richter, H.: Wahrscheinlichkeitstheorie. Springer, Berlin (1966)
41. Sawaragi, Y., Nakayama, H., Tanino, T.: Theory of Multiobjective Optimization. Academic, New York (1985)
42. Schoofs, A.: Experimental design and structural optimization. Ph.D. thesis, Mechanical Engineering (1987). https://doi.org/10.6100/IR270283
43. Schuëller, G., Gasser, M.: Some basic principles of reliability-based optimization (rbo) of structure and mechanical components. In: Marti, K., Kall, P. (eds.) Stochastic Programming Methods and Technical Applications. Lecture Notes in Economics and Mathematical Systems (LNEMS), vol. 458, pp. 80–103. Springer, Berlin (1998)
44. Schuëller, G., Stix, R.: A critical appraisal of methods to determine failure probabilities. Struct. Saf. **4**(4), 293–309 (1987). https://doi.org/10.1016/0167-4730(87)90004-X
45. Sciavicco, L., Siciliano, B.: Modeling and Control of Robot Manipulators. Springer, London (2000)
46. Slotine, J.J., Li, W.: Applied Nonlinear Control. Prentice-Hall Int. Inc., Englewood Chiffs (1991)
47. Stengel, R.: Stochastic Optimal Control: Theory and Application. Wiley, New York (1986)
48. Stöckl, G.: Optimaler Entwurf Elastoplastischer Mechanischer Strukturen Unter Stochastischer Unsicherheit. No. 278 in Fortschritt-Berichte VDI, Reihe 18. VDI-Verlag GmbH, Düsseldorf (2003)
49. Thoft-Christensen, P., Baker, M.: Structural Reliability Theory and its Applications. Springer, Berlin (1982)
50. Tong, Y.: Probability Inequalities in Multivariate Distributions. Academic, New York (1980)

Chapter 2
Solution of Stochastic Linear Programs by Discretization Methods

Abstract Solution procedures for stochastic linear optimization problems (also called stochastic linear programs (SLP)) by means of discretization of the probability distribution of the random parameters are treated in this chapter: Given a stochastic cost vector $c(\omega)$, a stochastic technology matrix $T(\omega)$ and a stochastic right-hand side, $h(\omega)$, consider a linear program for minimizing a linear function $c(\omega)^T x$ of the design vector x subject to the linear constraints $T(\omega)x = h(\omega)$, $x \geq 0$. Due to the stochastic variations of the data $(c, T, h) = (c(\omega), T(\omega), h(\omega))$, for the selection of an optimal decision vector x^*, an appropriate deterministic substitute problem has to be chosen. Here, we look for optimal decision vectors $x^* \geq 0$ minimizing the expected total cost defined by the sum of the primal costs $c(\omega)^T x$ and the costs $p(T(\omega)x - h(\omega))$ caused by the violation of the equality constraints $T(\omega)x = h(\omega)$. These costs are determined here by means of sublinear functions $p = p(z)$, involving, e.g., the class of norms for an error vector z. Moreover, several sublinear cost functions can be represented by the value function of an optimization problem, as, e.g., a Minkowski functional, see Chap. 11. Then, error estimates are given, and a priori bounds for the approximation error are derived. Furthermore, exploiting invariance properties of the probability distribution of the random parameters, problem-oriented discretizations are derived which simplify then the computation of admissible descent directions at non-stationary points.

Reproduced with permission from Springer Nature: Marti, K. (2002). On Solution of Stochastic Linear Programs by Discretization Methods. In: Dzemyda, G., Šaltenis, V., Žilinskas, A. (eds) Stochastic and Global Optimization. Nonconvex Optimization and Its Applications, vol 59. Springer, Boston, MA. https://doi.org/10.1007/0-306-47648-7_11.

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9_2

2.1 A Priori Error Bounds

A well-known method to handle linear programs

$$\begin{aligned} \min \quad & c(\omega)^T x \\ \text{s.t.} \quad & \\ & T(\omega)x = h(\omega), \\ & x \in D \end{aligned} \tag{2.1}$$

with random data $\Big(c(\omega), T(\omega), h(\omega)\Big)$ on a probability space $(\Omega, \mathcal{A}, P)$ is to replace (2.1), cf. [5, 10], by the stochastic optimization problem

$$\begin{aligned} \min \quad & F(x) \\ \text{s.t.} \quad & x \in D, \end{aligned} \tag{2.2a}$$

where

$$F(x) = E\Big(c(\omega)^T x + p\Big(h(\omega) - T(\omega)x\Big)\Big) \tag{2.2b}$$

and $p = p(z)$ denote the so-called second stage costs defined by

$$p(z) = \inf\{g^T y : Wy = z, y \geq 0\}, \quad z \in \mathbb{R}^m. \tag{2.2c}$$

Here, D is a fixed convex polyhedron in $\mathbb{R}^n$, hence, "$x \in D$" represents the deterministic constraints in (2.1), and "E" denotes the expectation operator. In the following we suppose that the $m \times n_1$ matrix W and the m vector q are related such that

$$\{Wy : x \geq 0\} = \mathbb{R}^m. \tag{2.3a}$$

$$\{v : W^T v \leq q\} \neq \emptyset. \tag{2.3b}$$

Thus, the loss function p is defined on the whole $\mathbb{R}^m$. If $q \geq 0$, then (2.3b) holds, and p is nonnegative. According to [4] we know that p is a sublinear function on $\mathbb{R}^m$, thus

$$p(z + w) \leq p(z) + p(w) \qquad \text{for all } z, w \in \mathbb{R}^m \tag{2.4a}$$

$$p(\lambda z) = \lambda p(z) \qquad \text{for all } z \in \mathbb{R}^m, \lambda \geq 0. \tag{2.4b}$$

Consequently, p is a convex function on $\mathbb{R}^m$, and we have that

$$p(0) = 0 \tag{2.4c}$$

$$-p(-z) \leq p(z) \qquad \text{for all } z \in \mathbb{R}^m \tag{2.4d}$$

$$-p(z - w) \leq p(w) - p(z) \leq p(w - z) \qquad \text{for all } z, w \in \mathbb{R}^m. \tag{2.4e}$$

Moreover, if

$$\|p\| = \sup_{\|z\|_E \le 1} \Big|p(z)\Big| = \sup_{\|z\|_E = 1} \Big|p(z)\Big| \ (< +\infty) \tag{2.4f}$$

denotes the norm of the sublinear function p, then

$$|p(z)| \le \|p\| \ \|z\| \qquad \text{for all } z \in \mathbb{R}^m, \tag{2.4g}$$

$$|p(w) - p(z)| \le \|p\| \ \|w - z\| \qquad \text{for all } z, w \in \mathbb{R}^m. \tag{2.4h}$$

Denoting $Ec(\omega)$ by $\overline{c}$, (2.2b) reads

$$F(x) = \overline{c}^T x + Ep\Big(h(\omega) - T(\omega)x\Big). \tag{2.5}$$

2.2 Disretization and Error Bounds

Approximating now the $m \times (n+1)$ random matrix $\Big(T(\omega), h(\omega)\Big)$ by a certain sequence of random matrices

$$\Big(T^1(\omega), h^1(\omega)\Big), \Big(T^2(\omega), h^2)(\omega)\Big), \ldots, \Big(T^N(\omega), h^N(\omega)\Big), \ldots,$$

converging in some probabilistic sense to $\Big(T(\omega), h(\omega)\Big)$, we obtain the approximative objective functions

$$F^N(x) = \overline{c}^T x + Ep\Big(h^N(\omega) - T^N(\omega)x\Big), \quad N = 1, 2, \ldots. \tag{2.6}$$

Using (2.4d), for $F(x) - F^N(x)$ we obtain, see [4].

$$-\vartheta^N(x) \le F(x) - F^N(x) \le \eta^N(x), \tag{2.7a}$$

where the lower, upper error term $\vartheta^N(x)$, $\eta^N(x)$, resp., is defined by

$$\vartheta^N(x) := Ep\left(\Big(T(\omega) - T^N(\omega)\Big)x + \Big(h^N(\omega) - h(\omega)\Big)\right) \tag{2.7b}$$

$$\eta^N(x) := Ep\left(\Big(T^N(\omega) - T(\omega)\Big)x + \Big(h(\omega) - h^N(\omega)\Big)\right). \tag{2.7c}$$

Using discretization methods, the approximations $\Big(T^N(\omega), h^N(\omega)\Big)$ are piecewise constant random variables, hence

$$\Big(T^N(\omega), h^N(\omega)\Big) = \Big(T^{N,j}, h^{N,j}\Big) \text{ for all } \omega \in \Omega^{N,j}, \ j = 1, 2, \ldots, r_N, \tag{2.8a}$$

where

$$\Omega^{N,1}, \Omega^{N,2}, \ldots, \Omega^{N,j}, \ldots, \Omega^{N,r_N} \text{ is a partition of } \Omega. \tag{2.8b}$$

Consequently, $F^N(x)$, given by (2.6), reads

$$F^N(x) = \overline{c}^T x + \sum_{j=1}^{r_N} P(\Omega^{N,j}) p(h^{N,j} - T^{N,j} x). \tag{2.8c}$$

Using Jensen's inequality, for the error estimate $\vartheta^N(x)$ we get

$$\begin{aligned}
\vartheta^N(x) &= \int p\left(\left(T(\omega) - T^N(\omega)\right)x + \left(h^N(\omega) - h(\omega)\right)\right) P(d\omega) \\
&= \sum_{j=1}^{r_N} \int_{\Omega^{N,j}} p\left(\left(T(\omega) - T^{N,j}\right)x + \left(h^{N,j} - h(\omega)\right)\right) P(d\omega) \\
&\geq \sum_{j=1}^{r_N} P(\Omega^{N,j}) p\left(\left(\overline{T}^{\Omega^{N,j}} - T^{N,j}\right)x + \left(n^{N,j} - \overline{h}^{\Omega^{N,j}}\right)\right),
\end{aligned} \tag{2.9a}$$

where

$$\overline{T}^{\Omega^{N,j}} = \frac{1}{P(\Omega^{N,j})} \int_{\Omega^{N,j}} T(\omega) P(d\omega), \tag{2.9b}$$

$$\overline{h}^{\Omega^{N,j}} = \frac{1}{P(\Omega^{N,j})} \int_{\Omega^{N,j}} h(\omega) P(d\omega) \tag{2.9c}$$

are the conditional expectation of $T(\omega)$, $h(\omega)$, resp., with respect to $\Omega^{N,j}$. Obviously, for $\eta^N(x)$ we obtain

$$\eta^N(x) \geq \sum_{j=1}^{r_N} P(\Omega^{N,j}) p\left(\left(T^{N,j} - \overline{T}^{\Omega^{N,j}}\right)x + \left(\overline{h}^{\Omega^{N,j}} - h^{N,j}\right)\right). \tag{2.9d}$$

Several publications [1, 2, 11] suggest to select the values $T^{N,j}$, $h^{N,j}$, resp., according to

$$T^{N,j} = \overline{T}^{\Omega^{N,j}} = \frac{1}{P(\Omega^{N,j})} \int_{\Omega^{N,j}} T(\omega) P(d\omega), \tag{2.10a}$$

$$h^{N,j} = \overline{h}^{\Omega^{N,j}} = \frac{1}{P(\Omega^{N,j})} \int_{\Omega^{N,j}} h(\omega) P(d\omega). \tag{2.10b}$$

In this case, (2.9a), (2.9d) and (2.4c) immediately yield this result:

Lemma 2.1 *If the approximation* $\left(T^N(\omega), h^N(\omega)\right)$ *of* $\left(T(\omega), h(\omega)\right)$ *is defined by* (2.8a), (2.8b) *and* (2.10a), (2.10b), *then* $\vartheta^N(x) \geq 0, \eta^N(x) \geq 0$ *for all* $x \in \mathbb{R}^n$.

If the norm $\|p\|$ *of the sublinear loss function is known, then the error terms* $\vartheta^N(x), \eta^N(x)$ *can be further estimated from above. Indeed,* (2.7b), (2.7c) *and* (2.4f) *yield*

$$|\vartheta^N(x)|, |\eta^N(x)| \leq \|p\| \left(E\left\|\left(T(\omega) - T^N(\omega)\right)x\right\| + E\left\|h^N(\omega) - h(\omega)\right\|\right). \tag{2.11a}$$

Denoting by

$$P^{N,j}(d\omega) = \frac{1}{P(\Omega^{N,j})} 1_{\Omega^{N,j}} P(d\omega) \tag{2.11b}$$

the restriction of P *to the subdomains* $\Omega^{N,j}$ *of* Ω, *according to* (2.8a)–(2.8c), *from* (2.11a), (2.11b) *we obtain*

$$\begin{aligned} &|\vartheta^N(x)|, |\eta^N(x)| \\ \leq &\|p\| \sum_{j=1}^{r_N} P(\Omega^{N,j}) \left(\int_{\Omega^{N,j}} \left\|\left(T(\omega) - T^{N,j}\right)x\right\| P^{N,j}(d\omega) \right. \\ &\left. + \int_{\Omega^{N,j}} \|h^{N,j} - h(\omega)\| P^{N,j}(d\omega) \right). \end{aligned} \tag{2.11c}$$

Remark 2.1 (*More general loss functions*) If the sublinear loss function p is replaced by a more general convex loss function u, then the inequalities (2.11a), (2.11c) remain true if the left-hand side in (2.11a), (2.11c) is simply replaced by $|F(x) - F^N(x)|$, and $\|p\|$ is replaced by a Lipschitz constant $L > 0$ of u, provided that u is Lipschitzian with constant L on the union of the supports of the random m-vectors $T(\omega)x - h(\omega)$ and $T^N(\omega)x - h^N(\omega),\ N = 1, 2, \ldots,\ x \in D$, cf. [4].

Because of

$$\begin{aligned} \left\|\left(T(\omega) - T^{N,j}\right)x\right\| &= \left(\left\|\left(T(\omega) - T^{N,j}\right)x\right\|^2\right)^{1/2} \\ &= \left(x^T\left(T(\omega) - T^{N,j}\right)^T\left(T(\omega) - T^{N,j}\right)x\right)^{1/2} \end{aligned}$$

and the concavity of $t \to \sqrt{t}$, we have that

$$\int_{\Omega^{N,j}} \left\| \left(T(\omega) - T^{N,j} \right) x \right\| P^{N,j}(d\omega) \leq \left(\int_{\Omega^{N,j}} \left\| \left(T(\omega) - T^{N,j} \right) x \right\|^2 P^{N,j}(d\omega) \right)^{1/2}$$

$$= \left(x^T \left(\sum_{i=1}^{m} \int_{\Omega^{N,j}} \left(T_i(\omega) - T_i^{N,j} \right)^T \left(T_i(\omega) - T_i^{N,j} \right) P^{N,j}(d\omega) \right) x \right)^{1/2}$$

$$\leq \|x\| \left(\sum_{i=1}^{m} \sum_{k=1}^{n} \int_{\Omega^{N,j}} \left(t_{ik}(\omega) - t_{ik}^{N,j} \right)^2 P^{N,j}(d\omega) \right)^{1/2} \tag{2.12a}$$

where $T_i, T_i^{N,j}$ is the i-th row of the $m \times n$ matrices $T(\omega) = \left(t_{ik}(\omega) \right)$, $T^{N,j} = \left(t_{ik}^{N,j} \right)$, respectively. Moreover,

$$\int_{\Omega^{N,j}} \|h^{N,j} - h(\omega)\| P^{N,j}(d\omega) \leq \left(\sum_{i=1}^{m} \int_{\Omega^{N,j}} \left(h_i^{N,j} - h_i(\omega) \right)^2 P^{N,j}(d\omega) \right). \tag{2.12b}$$

Clearly, if (2.10a), (2.10b) holds, then

$$\int_{\Omega^{N,j}} \left\| \left(T(\omega) - T^{N,j} \right) x \right\| P^{N,j}(d\omega) \leq \left(x' \left(\sum_{i=1}^{m} \operatorname{var}^{N,j} \left(T_i(\cdot) \right) \right) x \right)^{1/2} \tag{2.12c}$$

$$\int_{\Omega^{N,j}} \|h^{N,j} - h(\omega)\| P^{N,j}(d\omega) \leq \left(\sum_{i=1}^{m} \operatorname{var}^{N,j} \left(h_i(\cdot) \right) \right)^{1/2}, \tag{2.12d}$$

where $\operatorname{var}^{N,j} \left(T_i(\cdot) \right)$ is the covariance matrix of the i-th row $T_i(\omega)$ of $T(\omega)$, and $\operatorname{var}^{N,j} \left(h_i(\cdot) \right)$ is the variance of the i-th component $h_i(\omega)$ of $h(\omega)$ with respect to the conditional distribution $P^{N,j} = P|_{\Omega^{N,j}}$.

Remark 2.2 For the general case given by (2.12a), (2.12b) we have that

$$\int_{\Omega^{N,j}} \left(T_i(\omega) - T_i^{N,j} \right)^T \left(T_i(\omega) - T_i^{N,j} \right) P^{N,j}(d\omega) = \operatorname{var}^{N,j} \left(T_i(\cdot) \right) + \left(\overline{T}_i^{\Omega^{N,j}} - T_i^{N,j} \right)^T \left(\overline{T}^{\Omega^{N,j}} - T_i^{N,j} \right), \tag{2.12e}$$

$$\int_{\Omega^{N,j}} \left(h_i(\omega) - h_i^{N,j} \right)^2 P^{N,j}(d\omega) = \operatorname{var}^{N,j} \left(h_i(\cdot) \right) + \left(\overline{h}_i^{\Omega^{N,j}} - h_i^{N,j} \right)^2, \tag{2.12f}$$

respectively.

Because of (2.11c) and (2.12a), (2.12b) we still have to compute an upper estimate of $\|p\|$. Representing any vector $z \in \mathbb{R}^m$ by $z = \sum_{i=1}^m z_i e_i$, where $z_1, z_2, \ldots, z_m$ are the components of z and $e_1, e_2, \ldots, e_n$ are the unit vectors of the m coordinate directions, because of the sublinearity of p we find

$$\begin{aligned} p(z) &= p\left(\sum_{i=1}^m z_i e_i\right) \le \sum_{i=1}^m p(z_i e_i) = \sum_{i=1}^m p\Big((z_i^+ - z_i^-)e_i\Big) \\ &\le \sum_{i=1}^m \Big(p(z_i^+ e_i)\Big) + p\Big(z_i^-(-e_i)\Big) = \sum_{i=1}^m \Big(z_i^+ p(e_i) + z_i^- p(-e_i)\Big) \\ &\le \sum_{i=1}^m (z_i^+ + z_i^-) \max\Big\{p(e_i),\, p(-e_i)\Big\} = \sum_{i=1}^m \pi_i |z_i|, \end{aligned} \tag{2.13a}$$

where

$$\pi_i = \max\Big\{p(e_i),\, p(-e_i)\Big\}, \quad i = 1, 2, \ldots, m. \tag{2.13b}$$

Note that for the computation of the m coefficients $\pi_1, \pi_2, \ldots, \pi_m$ we have to solve the following $2m$ linear programs

$$\min q^T y \quad \text{s.t. } Wy = \pm e_i,\ y \ge 0 \quad \text{for} \quad i = 1, 2, \ldots, m. \tag{2.13c}$$

From (2.13a) we obtain

$$p(z) \le \|\pi\| \cdot \|z\|, \tag{2.13d}$$

where $\|\pi\|$, $\|z\|$ denote the Euclidean norm of $\pi = (\pi_1, \pi_2, \ldots, \pi_m)^T$ and z, respectively.

Thus, according to (2.4f) we have the upper norm bound

$$\|p\| \le \|\pi\| = \left(\sum_{i=1}^m \pi_i^2\right)^{1/2} = \left(\sum_{i=1}^m \Big(\max\Big\{p(e_i),\, p(-e_i)\Big\}\Big)^2\right)^{1/2}, \tag{2.14}$$

where $\|\pi\|$ **can be calculated from the given data**, cf. (2.2c), (2.13c).

Summarizing the above consierations, from (2.11c), (2.12a)–(2.12f), and (2.14) we get

Theorem 2.1 *If* $\Big(T^N(\omega), h^N(\omega)\Big)$ *is given by* (2.8a), (2.8b) *and* (2.3a), (2.3b) *holds, then for each* $x \in \mathbb{R}^n$ *we have that*

$$0 \le \vartheta^N(x), \eta^N(x) \le \|\pi\| \sum_{j=1}^{r_N} P(\Omega^{N,j}) \left(\|x\| \left(\sum_{i,k=1}^{m} \int_{\Omega^{N,j}} \left(t_{ik}(\omega) - t_{ik}^{N,j} \right)^2 P^{N,j}(d\omega) \right) \right)^{1/2} + \left(\sum_{i=1}^{m} \int_{\Omega_{N,j}} \left(h_i(\omega) - h_i^{N,j} \right)^2 P^{N,j}(d\omega) \right) \quad (2.15a)$$

Moreover, if equations (2.10a)–(2.10b) *hold, then*

$$0 \le \vartheta^N(x), \eta^N(x) \le \|\pi\| \sum_{j=1}^{r_N} P(\Omega^{N,j}) \left(\|x\| \left(\sum_{i,k=1}^{m} \operatorname{var}^{N,j} \left(t_{ik}(\cdot) \right) \right)^{1/2} + \left(\sum_{i=1}^{m} \operatorname{var}^{N,j} \left(h_i(\cdot) \right) \right)^{1/2} \right). \quad (2.15b)$$

2.2.1 Special Representations of the Random Matrix $\left(T(\cdot), h(\cdot)\right)$

A common, well-known representation of $\left(T(\omega), h(\omega)\right)$ is given by

$$\left(T(\omega), h(\omega)\right) = \left(T\left(\xi(\omega)\right), h\left(\xi(\omega)\right)\right) = (\overline{T}, \overline{h}) + \sum_{s=1}^{L} \xi_s(\omega) \left(T^{(s)}, h^{(s)}\right), \quad (2.16)$$

where $(\overline{T}, \overline{h})$ denotes the mean of $\left(T(\omega), h(\omega)\right)$, $\left(T^{(s)}, h^{(s)}\right)$, $s = 1, 2, \ldots, L$, are given $m \times (n+1)$ matrices, and $\xi_1(\omega), \xi_2(\omega), \ldots, \xi_L$, are zero-mean, stochastically independent random variables.

Based on representation (2.16), the approximation $\left(T^N(\omega), h^N(\omega)\right)$ of $\left(T(\omega), h(\omega)\right)$ according to (2.8a), (2.8b) can be described then by the following piecewise constant approximation $\xi^N(\omega)$ of the random L-vector $\xi(\omega) = \left(\xi_1(\omega), \xi_2(\omega), \ldots, \xi_L(\omega)\right)$.

Let Ξ denote the support of $\xi(\omega)$ or a set containing the support of $\xi(\omega)$. Moreover, let

$$\Xi^{N,1}, \Xi^{N,2}, \ldots, \Xi^{N,j}, \ldots, \Xi^{N,r_N} \quad (2.17a)$$

be the partition of Ξ generated by the partition (2.8b) of Ω, hence

$$\Xi^{N,j} = \left\{\xi(\omega) : \omega \in \Omega^{N,j}\right\}. \tag{2.17b}$$

Thus, with fixed L-vectors $\xi^{N,j} = \left(\xi_1^{N,j}, \ldots, \xi_L^{N,j}\right) \in \Xi^{N,j}$ we have that

$$\xi^N(\omega) = \xi^{N,j} \text{ for all } \omega \in \Omega^{N,j} \tag{2.17c}$$

and therefore

$$(T^{N,j}, h^{N,j}) = (\overline{T}, \overline{h}) + \sum_{s=1}^{L} \xi_s^{N,j}\left(T^{(s)}, h^{(s)}\right). \tag{2.17d}$$

Because of the properties of the random variables $\xi_1(\omega), \ldots, \xi_L(\omega)$ in the representation (2.16) of $\left(T(\omega), h(\omega)\right)$, we suppose that $E\xi_s^N(\omega) = 0$ and $E\xi_s^N(\omega)\xi_t^N(\omega) = 0$ for all $s = 1, \ldots, L, t = 1, \ldots, L, t \neq s$, hence

$$\sum_{j=1}^{r_N} \xi_s^{N,j} P_{\xi(\cdot)}(\Xi^{N,j}) = 0, \quad \sum_{j=1}^{r_N} \xi_x^{N,j}\xi_t^{N,j} P_{\xi(\cdot)}\left(\Xi^{N,j}\right) = 0, \tag{2.17e}$$

whereas $s, t = 1, \ldots, L, t \neq s$.

In many cases we may suppose that

$$\Xi = \prod_{s=1}^{L} \Xi_s,\ \Xi_s = [\alpha_s, \beta_s) \tag{2.18a}$$

is a half-open L-dimensional interval. In this case Ξ is then partitioned into certain subintervals $\Xi^{N,j}$. Hence, we set

$$j = (j_1, j_2, \ldots, j_L), r_N = (r_{N1}, r_{N2}, \ldots, r_{NL}), \tag{2.18b}$$

$$\xi^{N,j} = \left(\xi_1^{N,j_1}k, \xi_2^{N,j_2}, \ldots, \xi_L^{N,j_L}\right), \tag{2.18c}$$

where $j_s = 1, 2, \ldots, r_{N^s}, s = 1, \ldots, L$, and a cell $\Xi^{N,j}$ is given by

$$\Xi^{N,j} = \Xi^{N,(j_1,\ldots,j_L)} = \prod_{s=1}^{L} \Xi_s^{N,j_s} \tag{2.18d}$$

with certain half-open subintervals

$$\Xi_s^{N,j_s} = \left[\alpha_s^{N,j_s}, \beta_s^{N,j_s}\right), \quad j_s = 1, 2, \ldots, r_{N^s}, \quad s = 1, 2, \ldots, L \tag{2.18e}$$

with

$$\xi_s^{N,j_s} \in \left[\alpha_s^{N,j_s}, \beta_s^{N,j_s}\right), \quad j_s = 1, 2, \ldots, r_{N^s}, \quad s = 1, 2, \ldots, L. \tag{2.18f}$$

Moreover, (2.17d) reads in the present case

$$\begin{aligned}(T^{N,j}, h^{N,j}) &= \left(T^{N,(j_1,\ldots,j_L)}, h^{N,(j_1,\ldots,j_L)}\right) = \\ &= (\overline{T}, \overline{h}) = \sum_{s=1}^{L} \xi_s^{N,j_s}\left(T^{(s)}, h^{(s)}\right).\end{aligned} \tag{2.18g}$$

For the integrals in the error estimation (2.15a), by (2.16), (2.17d), (2.18g) and the cell representation (2.18d), we get

$$\int\limits_{\Omega^{N,j}} \left(t_{ik}(\omega) - t_{ik}^{N,j}\right)^2 P^{N,j}(d\omega) = \frac{1}{P_{\xi(\cdot)}(\Xi^{N,j})} \int\limits_{\xi \in \Xi^{N,j}} \left(t_{ik}(\xi) - t_{ik}^{N,j}\right)^2 P_{\xi(\cdot)}(d\xi)$$

and therefore

$$\begin{aligned}\int\limits_{\Omega^{N,j}} \left(t_{ik}(\omega) - t_{ik}^{N,j}\right)^2 P^{N,j}(d\omega) &= \sum_{\substack{s,\sigma=1 \\ s\neq\sigma}}^{L} t_{ik}^{(s)} t_{ik}^{(\sigma)} \left(\overline{\xi}_s^{\Xi_s^{N,j_s}} - \xi_s^{N,j_s}\right)\left(\overline{\xi}_\sigma^{\Xi_\sigma^{N,j_\sigma}} - \xi_\sigma^{N,j_\sigma}\right) \\ &+ \sum_{s=1}^{L} t_{ik}^{(s)2} \frac{1}{P_{\xi_s(\cdot)}(\Xi_s^{N,j_s})} \int\limits_{\xi_s \in \Xi_s^{N,j_s}} (\xi_s - \xi_s^{N,j_s})^2 P_{\xi_s(\cdot)}(d\xi_s)\end{aligned} \tag{2.19a}$$

as well as

$$\begin{aligned}\int\limits_{\Omega^{N,j}} \left(h_i(\omega) - h_i^{N,j}\right)^2 P^{N,j}(d\omega) &= \sum_{\substack{s,\sigma=1 \\ s\neq\sigma}}^{L} h_i^{(s)} h_i^{(\sigma)} \left(\overline{\xi}_s^{\Xi_s^{N,j_s}} - \xi_s^{N,j_s}\right)\left(\overline{\xi}_\sigma^{\Xi_\sigma^{N,j_\sigma}} - \xi_\sigma^{N,j_\sigma}\right) \\ &+ \sum_{s=1}^{L} h_i^{(s)2} \frac{1}{P_{\xi_s(\cdot)}(\Xi_s^{N,j_s})} \int\limits_{\xi_s \in \Xi_s^{N,j_s}} (\xi_s - \xi_s^{N,j_s})^2 P_{\xi_s(\cdot)}(d\xi_s),\end{aligned} \tag{2.19b}$$

where

$$\overline{\xi}_s^{\Xi_s^{N,j_s}} = E(\xi_s | \Xi_s^{N,j_s}) := \frac{1}{P_{\xi_s(\cdot)}(\Xi_s^{N,j_s})} \int\limits_{\xi_s \in \Xi_s^{N,j_s}} \xi_s P_{\xi_s(\cdot)}(d\xi_s) \tag{2.19c}$$

in the conditional mean of $\xi_s(\omega)$ with respect to Ξ_s^{N,j_s}, cf. (2.9b), (2.9c).

Since $\xi_s^{N,j_s} \in \left[\alpha_s^{N,j_s}, \beta_s^{N,j_s}\right)$, cf. (2.18f), and $\overline{\xi}_s^{\Xi_s^{N,j_s}} \in \left[\alpha_s^{N,j_s}, \beta_s^{N,j_s}\right)$, the first terms in (2.19a), (2.19b) can be estimated from above as follows:

$$\left|\sum_{\substack{s,\sigma=1\\ s\neq\sigma}}^{L} t_{ik}^{(s)} t_{ik}^{(\sigma)} \left(\overline{\xi}_s^{\Xi_s^{N,j_s}} - \xi_s^{N,j_s}\right)\left(\overline{\xi}_\sigma^{\Xi_\sigma^{N,j_\sigma}} - \xi_\sigma^{N,j_\sigma}\right)\right| \leq$$

$$\sum_{\substack{s,\sigma=1\\ s\neq\sigma}}^{L} \left|t_{ik}^{(s)} t_{ik}^{(\sigma)}\right| \left(\beta_s^{N,j_s} - \alpha_s^{N,j_s}\right)\left(\beta_\sigma^{N,j_\sigma} - \alpha_\sigma^{N,j_\sigma}\right), \tag{2.19d}$$

$$\left|\sum_{s,\sigma=1}^{L} h_i^{(s)} h_i^{(\sigma)} \left(\overline{\xi}_s^{\Xi_s^{N,j_s}} - \xi_s^{N,j_s}\right)\left(\overline{\xi}_\sigma^{\Xi_\sigma^{N,j_\sigma}} - \xi_\sigma^{N,j_\sigma}\right)\right| \leq$$

$$\sum_{\substack{s,\sigma=1\\ s\neq\sigma}}^{L} \left|h_i^{(s)} h_i^{(\sigma)}\right| \left(\beta_s^{N,j_s} - \alpha_s^{N,j_s}\right)\left(\beta_\sigma^{N,j_\sigma} - \alpha_\sigma^{N,j_\sigma}\right). \tag{2.19e}$$

If the interval $\Xi_s = [\alpha_s, \beta_s)$ is partitioned into r_{Ns} equidistant subintervals $[\alpha_s^{N,j_s}, \beta_s^{N,j_s})$, $j_s = 1, 2, \ldots, r_{Ns}$, then

$$\sum_{\substack{s,\sigma=1\\ s\neq\sigma}}^{L} \left|t_{ik}^{(s)} t_{ik}^{(\sigma)}\right| \left(\beta_s^{N,j_s} - \alpha_s^{N,j_s}\right)\left(\beta_\sigma^{N,j_\sigma} - \alpha_\sigma^{N,j_\sigma}\right)$$

$$= \frac{(\beta_s - \alpha_s)}{r_{Ns}} \frac{(\beta_\sigma - \alpha_\sigma)}{r_{N\sigma}} \sum_{\substack{s,\sigma=1\\ s\neq\sigma}}^{L} \left|t_{ik}^{(s)} t_{ik}^{(\sigma)}\right|, \tag{2.19f}$$

$$\sum_{\substack{s,\sigma=1\\ s\neq\sigma}}^{L} \left|h_i^{(s)} h_i^{(\sigma)}\right| \left(\beta_s^{N,j_s} - \alpha_s^{N,j_s}\right)\left(\beta_\sigma^{N,j_\sigma} - \alpha_\sigma^{N,j_\sigma}\right)$$

$$= \frac{(\beta_s - \alpha_s)^2}{r_{Ns} r_{N\sigma}} \sum_{\substack{s,\sigma=1\\ s\neq\sigma}}^{L} \left|h_i^{(s)} h_i^{(\sigma)}\right|. \tag{2.19g}$$

If the values ξ_s^{N,j_s}, $j_s = 1, 2, \ldots, r_{Ns}$, $s = 1, 2, \ldots, L$, are selected such that

$$\xi_s^{N,j_s} = \overline{\xi}_s^{\Xi_s^{N,j_s}}, \quad j_s = 1, 2, \ldots, r_N s, \quad s = 1, 2, \ldots, L, \tag{2.20a}$$

see (2.10a), (2.10b) then (2.19a), (2.19b) is reduced to

$$\int_{\Omega^{N,j}} \left(t_{ik}(\omega) - t_{ik}^{N,j}\right)^2 P^{N,j}(d\omega) = \sum_{s=1}^{L} t_{ik}^{(s)2} \operatorname{var}^{N,j_s}\left(\xi_s(\cdot)\right), \tag{2.20b}$$

$$\int_{\Omega^{N,j}} \left(h_i(\omega) - h_i^{N,j}\right)^2 P^{N,j}(d\omega) = \sum_{s=1}^{L} h_i^{(s)2} \operatorname{var}^{N,j_s}\left(\xi_s(\cdot)\right), \tag{2.20c}$$

where

$$\mathrm{var}^{N,j_s}\left(\xi_s(\cdot)\right) = \frac{1}{P_{\xi_s(\cdot)}(\Xi_s^{N,j_s})} \int\limits_{\xi_s \in \Xi_s^{N,j_s}} \left(\xi_s - \overline{\xi}_s^{\Xi_s^{N,j_s}}\right)^2 P_{\xi_s(\cdot)}(d\xi_s) \tag{2.20d}$$

is the conditional variance of $\xi_s(\omega)$ with respect to Ξ_s^{N,j_s}.

Example 2.1 Suppose that $\xi_s(\omega)$ has a density $f_s(z)$ such that for all $s = 1, 2, \ldots, L$ and $j_s = 1, 2, \ldots, r_{N,s}$ we have that

$$\begin{aligned} 0 < f_{s,m}^{N,j_s} &:= \inf\left\{f_s(z) : z \in \Xi_s^{N,j_s}\right\} \\ \leq f_{s,M}^{N,j_s} &:= \sup\left\{f_s(z) : z \in \Xi_s^{N,j_s}\right\} < +\infty. \end{aligned} \tag{2.21a}$$

This yields

$$P_{\xi_s(\cdot)}(\Xi_s^{N,j_s}) \geq f_{s,m}^{N,j_s}\left(\beta_s^{N,j_s} - \alpha_s^{N,j_s}\right), \tag{2.21b}$$

$$\int\limits_{\xi_s \in \Xi_s^{N,j_s}} \left(\xi_s - \overline{\xi}_s^{\Xi_s^{N,j_s}}\right)^2 P_{\xi_s(\cdot)}(d\xi_s) \leq f_{s,M}^{N,j_s} \int\limits_{\alpha_s^{N,j_s}}^{\beta_s^{N,j_s}} \left(\xi_s - \overline{\xi}_s^{\Xi_s^{N,j_s}}\right)^2 d\xi_s \tag{2.21c}$$

and therefore

$$\begin{aligned} \mathrm{var}^{N,j_s}\left(\xi_s(\cdot)\right) &\leq \frac{f_{s,M}^{N,j_s}}{f_{s,m}^{N,j_s}} \left(\frac{1}{\beta_s^{N,j_s} - \alpha_s^{N,j_s}} \int\limits_{\alpha_s^{N,j_s}}^{\beta_s^{N,j_s}} \left(\xi_s - \frac{\alpha_s^{N,j_s} + \beta_s^{N,j_s}}{2}\right)^2 d\xi_s \right. \\ &\qquad \left. + \left(\frac{\alpha_s^{N,j_s} + \beta_s^{N,j_s}}{2} - \overline{\xi}_s^{\Xi_s^{N,j_s}}\right)^2 \right) \\ &= \frac{f_{s,M}^{N,j_s}}{f_{s,m}^{N,j_s}} \left(\frac{(\beta_s^{N,j_s} - \alpha_s^{N,j_s})^2}{12} + \left(\frac{\alpha_s^{N,j_s} + \beta_s^{N,j_s}}{2} - \overline{\xi}_s^{\Xi_s^{N,j_s}}\right)^2 \right). \end{aligned} \tag{2.21d}$$

Since $\frac{\alpha_s^{N,j_s} + \beta_s^{N,j_s}}{2}$ and $\overline{\xi}_s^{\Xi_s^{N,j_s}}$ are elements of $\Xi_s^{N,j_s} = \left[\alpha_s^{N,j_s}, \beta_s^{N,j_s}\right)$, from (2.20d) and (2.21a)–(2.21d) we obtain

$$\mathrm{var}^{N,j_s}\left(\xi_s(\cdot)\right) \leq \frac{f_{s,M}^{N,j_s}}{f_{s,m}^{N,j_s}} \frac{13}{12} \left(\beta_s^{N,j_s} - \alpha_s^{N,j_s}\right)^2 \tag{2.22a}$$

If each Ξ_s is partitioned into r_{Ns} equidistant subintervals Ξ_s^{N,j_s}, $j_s = 1, 2, \ldots, r_{Ns}$, then $\beta_s^{N,j_s} - \alpha_s^{N,j_s} = \frac{1}{r_{Ns}}$ for all $j = 1, 2, \ldots, r_{Ns}$ and therefore

$$\mathrm{var}^{N,j_s}\left(\xi_s(\cdot)\right) \leq \frac{13}{12}\frac{f_{s,M}^{N,j_s}}{f_{s,m}^{N,j_s}} \cdot \frac{1}{r_{Ns}^2}. \tag{2.22b}$$

2.3 Approximations of F with a Given Error Level ε

According to (2.7a)–(2.7c) and (2.15a), (2.15b) we have that

$$\left|F(x) - F^N(x)\right| \leq \|\pi\| \sum_{j=1}^{r_N} P(\Omega^{N,j}) \left(\|x\| V^{N,j}\left(T(\cdot)\right) + V^{N,j}\left(h(\cdot)\right)\right), \tag{2.23a}$$

with the estimation errors

$$V^{N,j}\left(T(\cdot)\right) = \left(\sum_{i,k=1}^{N} \int_{\Omega^{N,j}} \left(t_{ik}(\omega) - t_{ik}^{N,j}\right)^2 P^{N,j}(d\omega)\right)^{1/2}, \tag{2.23b}$$

$$V^{N,j}\left(h(\cdot)\right) = \left(\sum_{i,k=1}^{N} \int_{\Omega^{N,j}} \left(h_i(\omega) - h_i^{N,j}\right)^2 P^{N,j}(d\omega)\right)^{1/2}. \tag{2.23c}$$

Knowing that D is bounded, hence

$$D \subset \left\{x \in \mathbb{R}^n : \|x\| \leq \rho_0\right\} \tag{2.24a}$$

for some $\rho_0 > 0$, from (2.23a) we obviously get

$$|F^* - F^{N^*}| \leq \|\pi\| \sum_{j=1}^{r_N} P(\Omega^{N,j}) \left(\rho_0 V^{N,j}\left(T(\cdot)\right) + V^{N,j}\left(h(\cdot)\right)\right), \tag{2.24b}$$

where F^* is the optimal value of (2.2a)–(2.2c) and F^{N^*} the optimal value of the approximating problem

$$\min F^N(x) \quad \text{s.t. } x \in D. \tag{2.24c}$$

Furthermore, if it is known that there is an optimal solution x^* of (2.2a)–(2.2c) such that with some $p \geq 1$ we have that

$$\|x^*\|_p \leq \rho_{0p} \text{ for some given } \rho_{0p} > 0, \tag{2.25a}$$

where $\|x\|_p$ is the p-norm of x, then again (2.7a)–(2.7c) and (2.15a), (2.15b) yield

$$\left|F^* - F_{\rho_0}^{N*}\right| \leq \|\pi\| \sum_{j=1}^{r_N} P(\Omega^{N,j}) \left(\rho_0 V^{N,j}\left(T(\cdot)\right) + V^{N,j}\left(h(\cdot)\right)\right), \tag{2.25b}$$

where $\rho_0 := \rho_{0p} \max\{\|u\| : \|u\|_p \leq 1\}$, and $F_{\rho_0}^{N*}$ is the optimal value of the approximation

$$\min F^N(x) \quad \text{s.t. } x \in D,\ \|x\|_p \leq \rho_{0p}. \tag{2.25c}$$

Note that if $F^N(x)$ is generated by a discretization process of $P_{\left(T(\cdot),h(\cdot)\right)}$, and $p = 1$ or $p = 1$ or $p = +\infty$, then (2.25c) can again be represented by a **linear program**.

The above considerations yield now the following result:

Theorem 2.2 *Selecting the approximation* $\left(T^N(\omega), h^N(\omega)\right)$ *of* $\left(T(\omega), h(\omega)\right)$ *such that*

$$\|\pi\| \sum_{j=1}^{r_N} P(\Omega^{N,j}) \left(\rho_0 V^{N,j}\left(T(\cdot)\right) + V^{N,j}\left(h(\cdot)\right)\right) \leq \varepsilon, \tag{2.26a}$$

where $\varepsilon > 0$ *is an a priori given error bound, then in cases* (2.24a) *and* (2.25a) *we have the* **a priori error bound**

$$|F^* - F^{N*}| < \varepsilon,\ |F^* - F_{\rho_0}^{N*}| < \varepsilon, \tag{2.26b}$$

respectively.

While (2.24a) is a simple property of the convex polyhedron D which may hold or not, the relation (2.25a) is more involved.

2.4 Norm Bounds for Optimal Solutions of (2.2a)–(2.2c)

For finding upper norm bounds ρ_0 for an optimal solution x^* of (2.2a)–(2.2c) we have to study the growth properties of F first. These can be obtained if for the loss function p, see (2.2c), appropriate lower bounds can be derived.

Using condition (2.3a), (2.3b), where we assume that q has components

$$q_1 > 0, q_2 > 0, \ldots, q_\mu > 0, \tag{2.27a}$$

we define now the closed, convex polyhedron K by

$$K = \text{conv} \left\{\frac{1}{q_1} w_1, \frac{1}{q_2} w_2, \ldots, \frac{1}{q_\mu} w_\mu\right\}, \tag{2.27b}$$

where $q_k > 0, k = 1, \ldots, \mu$, are the components of q and $w_1, w_2, \ldots, w_\mu$ are the columns of the matrix W, cf. (2.13c), where we may assume—without any restrictions—that $w_k \neq 0$ for all $k = 1, \ldots, \mu$. According to [3, 4] we know that in this situation the loss function p has the representation

$$p(z) = \inf \left\{ \lambda > 0 : \frac{z}{\lambda} \in K \right\}, z \in \mathbb{R}^m. \tag{2.28}$$

Having (2.27a), (2.27b) and defining

$$q_0 = \max_{1 \leq k \leq \mu} \frac{1}{q_k} \|w_k\|, \tag{2.29a}$$

where $q_0 > 0$, we find

$$\|\underline{z}\| \leq q_0 \quad \text{for each} \quad z \in K.$$

Since the relation $\frac{1}{\lambda} z \in K$ obviously implies that $\left\| \frac{1}{\lambda} z \right\| \leq q_0$, for the loss function p, having representation (2.28), for each $z \in \mathbb{R}^m$ we have that

$$p(z) \geq \inf \left\{ \lambda > 0 : \left\| \frac{z}{\lambda} \right\| \leq q_0 \right\} = \frac{1}{q_0} \|z\| = \left(\min_{1 \leq k \leq \mu} \frac{q_k}{\|w_k\|} \right) \|z\|. \tag{2.29b}$$

If (2.29b) holds, then (2.5) yields

$$F(x) \geq \overline{c}^T x + \underline{p} E \|T(\omega)x - h(\omega)\|, \tag{2.30a}$$

where

$$\underline{p} = \min_{1 \leq k \leq \mu} \frac{q_k}{\|w_k\|}. \tag{2.30b}$$

In the following we assume that $\Big(T(\omega), h(\omega)\Big)$ is bounded a.s., hence, there is a constant $\Gamma > 0$ such that

$$\left\| \Big(T(\omega), h(\omega)\Big) \right\| \leq \Gamma \text{ w.p. } 1. \tag{2.31}$$

Defining for any $x \in \mathbb{R}^n$

$$\hat{x} = (x^T, 1)^T, \quad e_{\hat{x}} = \frac{\hat{x}}{\|\hat{x}\|},$$

we find

$$\|T(\omega)x - h(\omega)\| = \left\| \Big(T(\omega), h(\omega)\Big) \hat{x} \right\| = \|\hat{x}\| \cdot \left\| \Big(T(\omega), h(\omega)\Big) e_{\hat{x}} \right\|;$$

furthermore, we have that

$$\begin{aligned}\left\|\Big(T(\omega),h(\omega)\Big)e_{\hat{x}}\right\|^2 &= \left\|\Big(T(\omega),h(\omega)\Big)e_{\hat{x}}\right\|\cdot\left\|\Big(T(\omega),h(\omega)\Big)e_{\hat{x}}\right\| \\ &\le \left\|\Big(T(\omega),h(\omega)\Big)e_{\hat{x}}\right\|\cdot\Gamma\end{aligned}$$

and therefore

$$\left\|T(\omega)x-h(\omega)\right\| \le \frac{1}{\Gamma}\|\hat{x}\|\cdot\left\|\Big(T(\omega),h(\omega)\Big)e_{\hat{x}}\right\|^2, \tag{2.32a}$$

see (2.31). Taking expectations on both sides of (2.32a), we get

$$E\|T(\omega)x-h(\omega)\| \ge \frac{1}{\Gamma}\|\hat{x}\|e_{\hat{x}}^T E\Big(T(\omega),h(\omega)\Big)^T\Big(T(\omega),h(\omega)\Big)e_{\hat{x}}. \tag{2.32b}$$

Denoting by $\lambda_{\min}(Q)$ the minimal eigenvalue of any symmetric matrix Q, from (2.32b) we obtain

$$\begin{aligned}E\|T(\omega)x-h(\omega)\| &\ge \frac{1}{\Gamma}\|\hat{x}\|\lambda_{\min}\left(E\Big(T(\omega),h(\omega)\Big)^T\Big(T(\omega),h(\omega)\Big)\right) \\ &\ge \frac{1}{\Gamma}\|\hat{x}\|E\lambda_{\min}\left(\Big(T(\omega),h(\omega)\Big)^T\Big(T(\omega),h(\omega)\Big)\right).\end{aligned} \tag{2.32c}$$

Note that

$$E\Big(T(\omega),h(\omega)\Big)^T\Big(T(\omega),h(\omega)\Big) = \sum_{i=1}^{m}\left(\operatorname{cov}\Big(T_i(\cdot),h_i(\cdot)\Big)+(\overline{T}_i,\overline{h}_i)^T(\overline{T}_i,\overline{h}_i)\right), \tag{2.32d}$$

where $\operatorname{cov}\Big(T_i(\cdot),h_i(\cdot)\Big)$ designates the covariance matrix of the i-th row (T_i,h_i) of $\Big(T(\omega),h(\omega)\Big)$. If the random matrix $\Big(T(\omega),h(\omega)\Big)$ is represented by (2.16), then

$$E\Big(T(\omega),h(\omega)\Big)^T\Big(T(\omega),h(\omega)\Big) = \overline{T}^T\overline{T}+\sum_{s=1}^{L}\operatorname{cov}(\xi_s){T^{(s)}}^T T^{(s)}, \tag{2.32e}$$

hence, this matrix can be computed easily. Let then λ_0 be defined by

$$\begin{aligned}\lambda_0 &:= \lambda_{\min}\left(E\Big(T(\omega),h(\omega)\Big)^T\Big(T(\omega),h(\omega)\Big)\right) \text{ or} \\ \lambda_0 &= E\lambda_{\min}\Big(T(\omega),h(\omega)\Big)^T\Big(T(\omega),h(\omega)\Big).\end{aligned} \tag{2.33}$$

Summarizing the above considerations, according to (2.30a) and (2.32b) we find the following result.

Lemma 2.2 *Suppose that conditions* (2.3a), (2.3b) *and* (2.31) *hold true. If* $\underline{p}$, Γ, λ_0 *are defined by* (2.30b), (2.31), (2.33), *resp., then*

$$F(x) \geq \overline{c}^T x + \frac{\underline{p}\lambda_0}{\Gamma}\|\hat{x}\| \quad \textit{for all} \;\; x \in \mathbb{R}^n. \tag{2.34}$$

Consider now any element x^0 of the feasible domain D of (2.2a). If x^* is an optimal solution of (2.2a)–(2.2c), then (2.34) and (2.13d) yield the following inequalities

$$\begin{aligned}\overline{c}^T x^* + \frac{\underline{p}\lambda_0}{\Gamma}\|\widehat{x^*}\| \leq F(x^*) \leq F(x^0) &= \overline{c}^T x^0 + \|\pi\|\,\|E\,|T(\omega)x^0 - h(\omega)\| \\ &\leq \overline{c}^T x^0 + \|\pi\|\left(\widehat{x^0}E\Big(T(\omega), h(\omega)\Big)^T\Big(T(\omega), h(\omega)\Big)\widehat{x^0}\right)^{1/2} \\ &:= F^0, \end{aligned} \tag{2.35}$$

where the last inequality is guaranteed by the concavity of the function $z \to \sqrt{z}$. Note that the upper bound F^0 can be computed easily, see (2.14), (2.32e). Since $\|\widehat{x^*}\| = \left(1 + \|x^*\|^2\right)^{1/2}$, we now have the following norm bounds for optimal solutions x^* of (2.2a)–(2.2c).

Theorem 2.3 *Suppose that the assumptions of Lemma 2.2 hold, and let* x^0 *be any feasible solution of* (2.2a)–(2.2c). *Moreover, let* F^0 *be defined as in* (2.35).

(a) If $\underline{c}'x \geq 0$ *for all* $x \in D$, *then for any optimal solution* x^* *of* (2.2a)–(2.2c) *we have that*

$$\|x^*\| \leq \left(\left(\frac{\Gamma}{\underline{p}\lambda_0 - \Gamma\|\underline{c}\|}F(x^0)\right)^2 - 1\right)^{1/2} \leq \left(\left(\frac{\Gamma F^0}{\underline{p}\lambda_0}\right)^2 - 1\right)^{1/2}. \tag{2.36a}$$

(b) If $\|\overline{c}\| < \dfrac{\underline{p}\lambda_0}{\Gamma}$, *then for any optimal solution* x^* *of* (2.2a)–(2.2c) *it holds*

$$\|x^*\| \leq \left(\left(\frac{F(x^0)\Gamma}{\underline{p}\lambda_0 - \Gamma\|\overline{c}\|}\right)^2 - 1\right)^{1/2} \leq \left(\left(\frac{F^0\Gamma}{\underline{p}\lambda_0 - \Gamma\|\overline{c}\|}\right)^2 - 1\right)^{1/2}. \tag{2.36b}$$

Proof

(a) Here, from (2.34) and (2.35) we get

$$F^0 \geq F(x^0) \geq \overline{c}^T x^* + \frac{p\lambda_0}{\Gamma}\|\widehat{x^*}\| \geq \frac{p\lambda_0}{\Gamma}\left(1 + \|x^*\|^2\right)^{1/2}$$

which yields the first assertion (2.36a).

(b) The next assertion (2.36b) follows from

$$\begin{aligned} F^0 \geq F(x^0) &\geq \overline{c}^T x^* + \frac{p\lambda_0}{\Gamma}\|\widehat{x^*}\| \geq -\|\overline{c}\| \cdot \|x^*\| + \frac{p\lambda_0}{\Gamma}\|\widehat{x^*}\| \\ &\geq \left(-\|\overline{c}\| + \frac{p\lambda_0}{\Gamma}\right)(1 + \|x^*\|^2)^{1/2}. \end{aligned}$$

□

2.5 Invariant Discretizations

According to Theorem 2.1, (2.23a)–(2.23c), there is a large variety of possible discretizations (T^N, h^N) of $\left(T(\omega), h(\omega)\right)$ guaranteeing a certain given a priori error bound, see (2.26b). Hence, the problem is to find discretizations taking into consideration the special structure of the underlying problem [6]. A main idea in stochastic linear programming with recourse is the use of special refining strategies for refining the partitions $\Xi^{N,1}, \ldots, \Xi^{N,r_N}$ of Ξ, see (2.17a)–(2.17e), such that only cells $\Xi^{N,j}$ are further partitioned which contribute most to the increase of the accuracy of approximation, see [1, 2].

Very often the probability distribution $P_{\left(T(\cdot),h(\cdot)\right)}$ has certain symmetry or invariance properties, [7]. Not destroying these invariance properties during the discretization process, in several cases descent discretions can be constructed very easily.

Considering the approximation $\left(T^N(\omega), h^N(\omega)\right)$, given by (2.8a), (2.8b) or (2.17a)–(2.17e), we define $\left(T_0^N(\omega), h_0^N(\omega)\right)$ by

$$\left(T_0^N(\omega), h_0^N(\omega)\right) := \left(T^N(\omega) - \overline{T}^N, h^N(\omega) - \overline{h}^N\right), \tag{2.37a}$$

where $\left(\overline{T}^N, \overline{h}^N\right)$ is the mean of $\left(T^N(\omega), h^N(\omega)\right)$. Using the results of [9], we define the distribution invariance as follows, where the set $\mathcal{B}_\alpha$ of $r_N \times r_N$ matrices $B = (b_{ij})$ is given by

$$\mathcal{B}_\alpha = \{B : 1^T B = 1^T, B\alpha = \alpha, B \geq 0\}. \tag{2.37b}$$

Here, 1 denotes the r_N-vector $1 = (1, 1, \ldots, 1)$ and α is the r_N-vector

$$\alpha = \left(P(\Omega^{N,1}),\, P(\Omega^{N,2}),\, \ldots,\, P(\Omega^{N,r_N})\right)^T \tag{2.37c}$$

or

$$\alpha = \left(P_{\xi(\cdot)}(\Xi^{N,1}),\, P_{\xi(\cdot)}(\Xi^{N,2}),\, \ldots,\, P_{\xi(\cdot)}(\Xi^{N,r_N})\right)^T \tag{2.37d}$$

and $B \geq 0$ means that $b_{ij} \geq 0$ for all elements b_{ij} of B.

Definition 2.1 The distribution $P_{\left(T_0^N(\cdot), h_0^N(\cdot)\right)}$ of $\left(T_0^N(\omega), h_0^N(\omega)\right)$ is called **invariant** if there is a matrix $B \in \mathcal{B}_\alpha$ and an $n \times n$ matrix C such that for each row $i = 1, 2, \ldots, m$ we have that

$$B^T \begin{pmatrix} T_{0,i}^{N,1} \\ T_{0,i}^{N,2} \\ \vdots \\ T_{0,i}^{N,r_N} \end{pmatrix} = \begin{pmatrix} T_{0,i}^{N,1} \\ T_{0,i}^{N,2} \\ \vdots \\ T_{0,i}^{N,r_N} \end{pmatrix} C \tag{2.38a}$$

$$B^T \begin{pmatrix} h_{0,i}^{N,1} \\ h_{0,i}^{N,2} \\ \vdots \\ T_{0,i}^{N,r_N} \end{pmatrix} = \begin{pmatrix} h_{0,i}^{N,1} \\ h_{0,i}^{N,2} \\ \vdots \\ h_{0,i}^{N,r_N} \end{pmatrix}. \tag{2.38b}$$

For the general case, we have to introduce some more notations:
Let denote $\tilde{z}$ the $(1 + m)$-vector

$$\tilde{z} = \begin{pmatrix} t \\ z \end{pmatrix} \quad \text{with} \quad t \in \mathbb{R},\ z \in \mathbb{R}^m, \tag{2.39a}$$

where we set $\tilde{z} = (\tilde{z}_0, \tilde{z}_1, \ldots, \tilde{z}_m)$ with $\tilde{z}_0 = t$, $\tilde{z}_i = z_i$, $i = 1, \ldots, m_i$. Furthermore, let $u(\tilde{z})$ denote the total loss function

$$u(\tilde{z}) = t + p(z) \tag{2.39b}$$

of (2.2a)–(2.2c).

Obviously, the total loss function $u = u(\tilde{z})$ is monotonous nondecreasing with respect to the component $z_0 = t$. In many cases the loss function p itself has some (partial) monotonicity properties, see, e.g., [8]. Hence, supposing in the following—for example—that p is also **partially nondecreasing**, we have a subset $J \subset \{0, 1, \ldots, m\}$ with $0 \in J$ and a corresponding partition

$$\tilde{z} = \begin{pmatrix} \tilde{z}_I \\ \tilde{z}_{II} \end{pmatrix}, \quad \tilde{z}_I = (\tilde{z}_i)_{i \in J}, \quad \tilde{z}_{II} = (z_i)_{i \notin J} \tag{2.40a}$$

of $\tilde{z}$ into two subvectors $\tilde{z}_I, \tilde{z}_{II}$ such that for any vectors $\tilde{z}, \tilde{w} \in \mathbb{R}^{1+m}$ the following relation hold:

$$\tilde{z}_I \leq \tilde{w}_I,\ \tilde{z}_{II} = \tilde{w}_{II} \quad \Rightarrow \quad u(\tilde{z}) \leq u(\tilde{w}), \tag{2.40b}$$

where $\tilde{z}_I \leq \tilde{w}_I$ means that $z_i \leq w_i$ for all $i \in J$. Of course, in many cases we also have the stronger condition

$$\tilde{z}_I \leq \tilde{w}_I,\ \tilde{z}_{II} = \tilde{w}_{II},\ \tilde{z}_i < \tilde{w}_i \ \text{ for at least one } \ i \in J \quad \Rightarrow \quad u(\tilde{z}) < u(\tilde{w}). \tag{2.40c}$$

Based on the above definitions, the invariance of an arbitrary distribution $P_{\left(A^N(\cdot), b^N(\cdot)\right)}$ with

$$\left(A^N(\omega), b^N(\omega)\right) = \begin{pmatrix} \overline{c}^T & 0 \\ T^N(\omega) & h^N(\omega) \end{pmatrix} \tag{2.41}$$

is stated as follows, where the following inclusion is still assumed:

$$D \subset \mathbb{R}^n_+. \tag{2.42}$$

Definition 2.2 The probability distribution $P_{\left(A^N(\cdot), b^N(\cdot)\right)}$ of $\left(A^N(\omega), b^N(\omega)\right)$ is called **invariant** it there is a matrix $B \in \mathcal{B}_\alpha$ and an $n \times n$ matrix C such that the following relations hold:

(i)

$$\overline{c} \geq \overline{c}^T C \tag{2.43a}$$

(ii)

$$\overline{T}_I \leq \overline{T}_I C \tag{2.43b}$$
$$\overline{T}_{II} = \overline{T}_{II} C \tag{2.43c}$$

(iii)

$$\text{(2.8a) and (2.38b) are fulfilled.} \tag{2.43d}$$

where $\overline{T}_I, \overline{T}_{II}$, resp. is the matrix containing the rows $\overline{T}_i$ with $i \in J, i \notin J$, respectively.

Remark 2.3 Note that condition (2.43d), hence, relations (2.38a) and (2.38b) can be interpreted as **conditions for the discretization** of the distribution of the centralized random matrix $\left(T_0(\omega), h_0(\omega)\right) = \left(T(\omega) - \overline{T}, h(\omega) - \overline{h}\right)$, where $(\overline{T}, \overline{h})$ is the mean of $\left(T(\omega), h(\omega)\right)$.

The significance of the above invariance concept follows from the following result, cf. [9].

Theorem 2.4 *Suppose that* $D \subset \mathbb{R}^n_+$. *If* $\left(A^N(\cdot), b^N(\cdot)\right)$ *has an invariant distribution with matrices* $B \in \mathcal{B}_\alpha$, C *according to Definition 2.2, then*

(I) $F^N(y) \leq F^N(x)$ *with* $y := Cx$ *for every* $x \in \mathbb{R}^n$

(II) $h = y - x$ *is a descent direction for* F^N *at* x, *provided that only* F^N *is not constant on the line segment* xy *joining* x *and* $y \neq x$.

As an important consequence of Theorem 5.1 we find the following result:

Corollary 2.1 *Assume that* $P_{\left(A^N(\cdot), b^N(\cdot)\right)}$ *is invariant with matrices* $B \in \mathcal{B}_\alpha$, C *according to Definition 2.2. Furthermore, suppose that* $F^N(x)$ *is not constant on each line segment* xy *in* D. *If* x^* *is an optimal solution of the approximating problem (2.24c), then*

$$Cx^* = x^* \text{ or}$$
$$h = Cx^* - x^* \text{ is not a feasible direction for } D \text{ at } x^*.$$

Note 2.1 Corollary 2.1 holds also under weaker conditions concerning F^N.

References

1. Kall, P.: Stochastic Linear Programming. Springer, Berlin (1976)
2. Kall, P., Wallace, S.: Stochastic Programming. Stochastic Programming. Wiley, Chichester (1994)
3. Marti, K.: Entscheidungsprobleme mit linearem Aktionen- und Ergebnisraum. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **23**, 133–147 (1972)
4. Marti, K.: Approximationen der Entscheidungsprobleme mit linearer Ergebnisfunktion und positiv homogener, subadditiver Verlustfunktion. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **31**, 203–233 (1975)
5. Marti, K.: Approximationen stochastischer Optimierungsprobleme. Hain Konigstein/Ts (1979)
6. Marti, K.: Diskretisierung stochastischer Programme unter Berücksichtigung der Problemstruktur. Z. Angew. Math. Mech. **59**, T105–T108 (1979)
7. Marti, K.: Computation of descent directions in stochastic optimization problems with invariant distributions. ZAMM - J. Appl. Math. Mech./Zeitschrift für Angewandte Mathematik und Mechanik **65**(8), 355–378 (1985). https://doi.org/10.1002/zamm.19850650813
8. Marti, K.: Descent Directions and Efficient Solutions in Discretely Distributed Stochastic Programs. Lecture Notes in Economics and Mathematical Systems, vol. 299. Springer, Berlin (1988)
9. Marti, K.: Computation of efficient solutions of discretely distributed stochastic optimization problems. ZOR Methods Model. Oper. Res. **36**(3), 259–294 (1992). https://doi.org/10.1007/BF01415892
10. Marti, K.: Optimal design of trusses as a stochastic linear programming problem. In: Nowak, A. (ed.) Reliability and Optimization of Structural Systems, pp. 231–239. University of Michigan Press, Ann Arbor (1998)
11. Mayer, J.: Stochastic Linear Programming Algorithms: A Comparison Based on a Model Management System. Gordon and Breach Science Publishers, Amsterdam (1998)

Chapter 3
Optimal Control Under Stochastic Uncertainty

Abstract Optimal control problems as arising in different technical (mechanical, electrical, thermodynamic, chemical, etc.) plants and economic systems are modeled mathematically by a system of first-order nonlinear differential equations for the plant state vector $z = z(t)$ involving, e.g., displacements, stresses, voltages, currents, pressures, concentration of chemicals, demands, etc. This system of differential equations depends on the vector $u(t)$ of input or control variables and a vector $a = a(\omega)$ of certain random model parameters. Moreover, also the vector z_0 of initial values of the plant state vector $z = z(t)$ at the initial time $t = t_0$ may be subject to random variations. While the actual realizations of the random parameters and initial values are not known at the planning stage, we may assume that the probability distribution or at least the occurring moments, such as expectations, variances, etc., are known. Moreover, we suppose that the costs along the trajectory and the terminal costs G are convex functions with respect to the pair (u, z) of control and state variables u, z, the final state $z(t_f)$, respectively. The problem is then to determine an open-loop, closed-loop, or an intermediate open-loop feedback control law minimizing the expected total costs consisting of the sum of the costs along the trajectory and the terminal costs. For the computation of stochastic optimal open-loop controls at each starting time point t_b, the stochastic Hamilton function of the control problem is introduced first. Then, a H-minimal control can be determined by solving a finite-dimensional stochastic optimization problem for minimizing the conditional expectation of the stochastic Hamiltonian subject to the remaining deterministic control constraints at each time point t. Having a H-minimal control, the related Hamiltonian two-point boundary value problem with random parameters is formulated for the computation of the stochastic optimal state and adjoint state trajectory. In the case of a linear-quadratic control problem the state and adjoint state trajectory can be determined analytically to a large extent. Inserting then these trajectories into the H-minimal control, stochastic optimal open-loop controls are found. For approximate solutions of the stochastic two-point boundary problem, cf. [31].

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9_3

3.1 Stochastic Control Systems

Optimal control and regulator problems arise in many concrete applications (mechanical, electrical, thermodynamical, chemical, etc.) are modeled [3, 17, 34, 37] by dynamical control systems obtained from physical measurements and/or known physical laws. The basic control system (input-output system) is mathematically represented [18, 38] by a system of first-order random differential equations:

$$\dot{z}(t) = g\Big(t, \omega, z(t), u(t)\Big), \quad t_0 \le t \le t_f, \ \omega \in \Omega \tag{3.1a}$$

$$z(t_0) = z_0(\omega). \tag{3.1b}$$

Here, ω is the basic random element taking values in a probability space $(\Omega, \mathfrak{A}, P)$, and describing the present random variations of model parameters or the influence of noise terms. The probability space $(\Omega, \mathfrak{A}, P)$ consists of the sample space or set of elementary events Ω, the σ-algebra $\mathfrak{A}$ of events and the probability measure P. The plant state vector $z = z(t, \omega)$ is an m-vector involving direct or indirect measurable/observable quantities like displacements, stresses, voltage, current, pressure, concentrations, etc., and their time derivatives (velocities), $z_0(\omega)$ is the random initial state. The plant control or control input $u(t)$ is a deterministic or stochastic n-vector denoting system inputs like external forces or moments, voltages, field current, thrust program, fuel consumption, production rate, etc. Furthermore, $\dot{z}$ denotes the derivative with respect to the time t. We assume that an input $u = u(t)$ is chosen such that $u(\cdot) \in U$, where U is a suitable linear space of input functions $u(\cdot) : [t_0, t_f] \to \mathbb{R}^n$ on the time interval $[t_0, t_f]$. Examples for U are subspaces of the space $PC_0^n[t_0, t_f]$ of piecewise continuous functions $u(\cdot) : [t_0, t_f] \to \mathbb{R}^n$ normed by the supremum norm

$$\Big\|u(\cdot)\Big\|_\infty = \sup\Big\{\big\|u(t)\big\| : t_0 \le t \le t_f\Big\}.$$

Note that a function on a closed, bounded interval is called *piecewise continuous* if it is continuous up to at most a finite number of points, where the one-sided limits of the function exist. Other important examples for U are the Banach spaces of integrable, essentially bounded measurable or regulated [8] functions $L_p^n([t_0, t_f], \mathcal{B}^1, \lambda^1)$, $p \ge 1$, $L_\infty^n([t_0, t_f], \mathcal{B}^1, \lambda^1)$, $Reg([t_0, t_f]; \mathbb{R}^n)$, resp., on $[t_0, t_f]$. Here, $\big([t_0, t_f], \mathcal{B}^1, \lambda^1\big)$ denotes the measure space on $[t_0, t_f]$ with the $\sigma-$ algebra $\mathcal{B}^1$ of Borel sets and the Lebesgue-measure λ^1 on $[t_0, t_f]$. Obviously, $PC_0^n[t_0, t_f] \subset L_\infty^n([t_0, t_f], \mathcal{B}^1, \lambda^1)$. If $u = u(t, \omega)$, $t_0 \le t \le t_f$, is a random input function, then correspondingly we suppose that $u(\cdot, \omega) \in U$ a.s. (almost sure or with probability 1). Moreover, we suppose that the function $g = g(t, \omega, z, u)$ of the plant differential equation (3.1a) and its partial derivatives (Jacobians) $D_z g$, $D_u g$ with respect to z and u are at least measurable on the space $[t_0, t_f] \times \Omega \times \mathbb{R}^m \times \mathbb{R}^n$.

The possible trajectories of the plant, hence, absolutely continuous [32] m-vector functions, are contained in the linear space $Z = C_0^m[t_0, t_f]$ of continuous functions $z(\cdot) : [t_0, t_f] \to \mathbb{R}^m$ on $[t_0, t_f]$. The space Z contains the set $PC_1^m[t_0, t_f]$ of con-

tinuous, piecewise differentiable functions on the interval $[t_0, t_f]$. A function on a closed, bounded interval is called *piecewise differentiable* if the function is differentiable up to at most a finite number of points, where the function and its derivative have existing one-sided limits. The space Z is also normed by the supremum norm. $D(\subset U)$ denotes the convex set of admissible controls $u(\cdot)$, defined, e.g., by some box constraints. Using the available information $\mathfrak{A}_t$ up to a certain time t, the problem is then to find an optimal control function $u^* = u^*(t)$ being most insensitive with respect to random parameter variations. This can be obtained by minimizing the total (conditional) expected costs arising along the trajectory $z = z(t)$ and/or at the terminal state $z_f = z(t_f)$ subject to the plant differential equation (3.1a), (3.1b) and the required control and state constraints. Optimal controls being most insensitive with respect to random parameter variations are also called *robust* controls. Such controls can be obtained by stochastic optimization methods [26].

Since feedback control (FB) laws can be approximated very efficiently, cf. [2, 19, 34], by means of *open-loop feedback (OLF)* control laws, see Sect. 3.2, for practical purposes we may confine to the computation of deterministic stochastic optimal open-loop (OL) controls $u = u(\cdot; t_b)$, $t_b \le t \le t_f$, on arbitrary "*remaining time intervals*" $[t_b, t_f]$ of $[t_0, t_f]$. Here, $u = u(\cdot; t_b)$ is stochastic optimal with respect to the information $\mathfrak{A}_{t_b}$ at the "initial" time point t_b.

3.1.1 Random Differential and Integral Equations

In many technical applications the random variations are not caused by an additive white noise term, but by means of possibly time-dependent random parameters. Hence, in the following the dynamics of the control system is represented by random differential equation, i.e., a *system of ordinary differential equations* (3.1a), (3.1b) *with random parameters*. Furthermore, solutions of random differential equations are defined here in the parameter (point)-wise sense, cf. [4, 6].

In case of a *discrete* or *discretized* probability distribution of the random elements, model parameters, i.e., $\Omega = \{\omega_1, \omega_2, \ldots, \omega_\varrho\}$, $P(\omega = \omega_j) = \alpha_j > 0$, $j = 1, \ldots, \varrho$, $\sum_{j=1}^{\varrho} \alpha_j = 1$, we can *redefine* (3.1a), (3.1b) by

$$\dot{z}(t) = g\Big(t, z(t), u(t)\Big), \quad t_0 \le t \le t_f, \tag{3.1c}$$

$$z(t_0) = z_0, \tag{3.1d}$$

with the vectors and vector functions

$$z(t) := \Big(z(t, \omega_j)\Big)_{j=1,\ldots,\varrho}, \qquad z_0 := \Big(z_0(\omega_j)\Big)_{j=1,\ldots,\varrho}$$

$$g(t, z, u) := \Big(g(t, \omega_j, z_{(j)}, u)\Big)_{j=1,\ldots,\varrho}, \qquad z := (z_{(j)})_{j=1,\ldots,\varrho} \in \mathbb{R}^{\varrho m}.$$

Hence, in this case (3.1c), (3.1d) represent again an ordinary system of first order differential equations for the ϱm unknown functions

$$z_{ij} = z_i(t, \omega_j), \quad i = 1, \ldots, m, \quad j = 1, \ldots, \varrho.$$

Results on the existence and uniqueness of the systems (3.1a), (3.1b) and (3.1c), (3.1d) and their dependence on the inputs can be found in [8].

Also in the general case we consider a *solution in the point-wise sense*. This means that for each random element $\omega \in \Omega$, (3.1a), (3.1b) is interpreted as a system of ordinary first-order differential equations with the initial values $z_0 = z_0(\omega)$ and control input $u = u(t)$. Hence, we assume that to each deterministic control $u(\cdot) \in U$ and each random element $\omega \in \Omega$ there exists a unique solution

$$z(\cdot, \omega) = S\Big(\omega, u(\cdot)\Big) = S\Big(\omega, u(\cdot)\Big), \tag{3.2a}$$

$z(\cdot, \omega) \in C_0^m[t_0, t_f]$, of the integral equation

$$z(t) = z_0(\omega) + \int_{t_0}^{t} g\Big(s, \omega, z(s), u(s)\Big) ds, \quad t_0 \le t \le t_f, \tag{3.2b}$$

such that $(t, \omega) \to S\Big(\omega, u(\cdot)\Big)(t)$ is measurable. This solution is also denoted by

$$z(t, \omega) = z_u(t, \omega) = z(t, \omega, u(\cdot)), \quad t_0 \le t \le t_f. \tag{3.2c}$$

Obviously, the integral equation (3.2b) is the integral version of the initial value problem (3.1a), (3.1b): Indeed, if, for given $\omega \in \Omega$, $z = z(t, \omega)$ is a solution of (3.1a), (3.1b), i.e., $z(\cdot, \omega)$ is absolutely continuous, satisfies (3.1a) for almost all $t \in [t_0, t_f]$ and fulfills (3.1b), then $z = z(t, \omega)$ is also a solution of (3.2b). Conversely, if, for given $\omega \in \Omega$, $z = z(t, \omega)$ is a solution of (3.2b), such that the integral on the right-hand side exists in the Lebesgue-sense for each $t \in [t_0, t_f]$, then this integral as a function of the upper bound t and therefore also the function $z = z(t, \omega)$ is absolutely continuous. Hence, by taking $t = t_0$ and by differentiation of (3.2b) with respect to t, cf. [32], we have that $z = z(t, \omega)$ is also a solution of (3.1a), (3.1b).

3.1.1.1 Parametric Representation of the Random Differential/Integral Equation

In the following we want to justify the above assumption that the initial value problem (3.1a), (3.1b), the equivalent integral equation (3.2b), resp., has a unique solution $z = z(t, \omega)$. For this purpose, let $\theta = \theta(t, \omega)$ be an r-dimensional stochastic process, as, e.g., time-varying disturbances, random parameters, etc., of the system, such that the sample functions $\theta(\cdot, \omega)$ are continuous with probability one. Furthermore, let

$$\tilde{g} : [t_0, t_f] \times \mathbb{R}^r \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$$

be a continuous function having continuous Jacobians $D_\theta \tilde{q}$, $D_z \tilde{g}$, $D_u \tilde{g}$ with respect to θ, z, u. Now consider the case that the function g of the process differential equation (3.1a), (3.1b) is given by

$$g(t, \omega, z, u) := \tilde{g}\Big(t, \theta(t, \omega), z, u\Big), \quad (t, \omega, z, u) \in [t_0, t_f] \times \Omega \times \mathbb{R}^m \times \mathbb{R}^n.$$

The spaces U, Z of possible inputs, trajectories, resp., of the plant are chosen as follows: $U := Reg([t_0, t_f]; \mathbb{R}^n)$ is the Banach space of all regulated functions $u(\cdot) : [t_0, t_f] \to \mathbb{R}^n$, normed by the supremum norm $\|\cdot\|_\infty$.

Furthermore, we set $Z := C_0^m[t_0, t_f]$ and $\Theta := C_0^r[t_0, t_f]$. Here, for an integer ν, $C_0^\nu[t_0, t_f]$ denotes the Banach space of all continuous functions of $[t_0, t_f]$ into $\mathbb{R}^\nu$ normed by the supremum norm $\|\cdot\|_\infty$. By our assumption we have $\theta(\cdot, \omega) \in \Theta$ a.s. (almost sure). Define

$$\Xi = \mathbb{R}^m \times \Theta \times U;$$

Ξ is the space of possible *initial values, time-varying model/environmental parameters and inputs* of the dynamic system. Hence, Ξ may be considered as the total space of inputs

$$\xi := \begin{pmatrix} z_0 \\ \theta(\cdot) \\ u(\cdot) \end{pmatrix}$$

into the plant, consisting of the random initial state z_0, the random input function $\theta = \theta(t, \omega)$ and the control function $u = u(t)$. Let now the mapping $\tau : \Xi \times Z \to Z$ related to the plant equation (3.1a), (3.1b) or (3.2b) be given by

$$\tau\Big(\xi, z(\cdot)\Big)(t) = z(t) - \left(z_0 + \int_{t_0}^{t} \tilde{g}\Big(s, \theta(s), z(s), u(s)\Big) ds\right), \quad t_0 \le t \le t_f. \tag{3.2d}$$

Note that for each input vector $\xi \in \Xi$ and function $z(\cdot) \in Z$ the integrand in (3.2d) is piecewise continuous, bounded or at least essentially bounded on $[t_0, t_f]$. Hence, the integral in (3.2d) as a function of its upper limit t yields again a continuous function on the interval $[t_0, t_f]$, and therefore an element of Z. This shows that τ maps $\Xi \times Z$ into Z.

Obviously, the initial value problem (3.1a), (3.1b) or its integral form (3.2b) can be represented by the operator equation

$$\tau(\xi, z(\cdot)) = 0. \tag{3.2e}$$

Operators of the type (3.2d) are well studied, see, e.g., [8, 20]: It is known that τ is continuously Fréchet (F)-differentiable [8, 20]. Note that the F-differential is a generalization of the derivatives (Jacobians) of mappings between finite-dimensional spaces to mappings between arbitrary normed spaces. Thus, the F-derivative $D\tau$ of τ at a certain point $\left(\bar{\xi}, \bar{z}(\cdot)\right)$ is given by

$$\begin{aligned}
&\left(D\tau\left(\bar{\xi}, \bar{z}(\cdot)\right) \cdot \left(\xi, z(\cdot)\right)\right)(t) \qquad (3.2g)\\
&= z(t) - \left(z_0 + \int_{t_0}^{t} D_z \tilde{g}\left(s, \bar{\theta}(s), \bar{z}(s), \bar{u}(s)\right) z(s) ds\right.\\
&+ \int_{t_0}^{t} D_\theta \tilde{g}\left(s, \bar{\theta}(s), \bar{z}(s), \bar{u}(s)\right) \theta(s) ds\\
&\left.+ \int_{t_0}^{t} D_u \tilde{g}\left(s, \bar{\theta}(s), \bar{z}(s), \bar{u}(s)\right) u(s) ds\right), \; t_0 \le t \le t_f,
\end{aligned}$$

where $\bar{\xi} = \left(\bar{z}_0, \bar{\theta}(\cdot), \bar{u}(\cdot)\right)$ and $\xi = \left(z_0, \theta(\cdot), u(\cdot)\right)$. Especially, for the derivative of τ with respect to $z(\cdot)$ we find

$$\left(D_z\tau\left(\bar{\xi}, \bar{z}(\cdot)\right) \cdot z(\cdot)\right)(t) = z(t) - \int_{t_0}^{t} D_z \tilde{g}\left(s, \bar{\theta}(s), \bar{z}(s), \bar{u}(s)\right) z(s) ds, \; t_0 \le t \le t_f. \qquad (3.2g)$$

The related equation

$$D_z\tau\left(\bar{\xi}, \bar{z}(\cdot)\right) \cdot z(\cdot) = y(\cdot), \; y(\cdot) \in Z \qquad (3.2h)$$

is a linear vectorial Volterra integral equation. By our assumptions this equation has a unique solution $z(\cdot) \in Z$. Note, that the corresponding result for scalar Volterra equations, see, e.g., [36], can be transferred to the present vectorial case. Therefore, $D_z\tau\left(\bar{\xi}, \bar{z}(\cdot)\right)$ is a linear, continuous one-to-one map from Z onto Z.

Hence, its inverse $\left(D_z\tau\left(\bar{\xi}, \bar{z}(\cdot)\right)\right)^{-1}$ exists. Using the implicit function theorem [8, 20], we now obtain the following result:

Lemma 3.1 *For given $\bar{\xi} = \left(\bar{z}_0, \bar{\theta}(\cdot), \bar{u}(\cdot)\right)$, let $\left(\bar{\xi}, \bar{z}(\cdot)\right) \in \Xi \times Z$ be selected such that $\tau\left(\bar{\xi}, \bar{z}(\cdot)\right) = 0$, hence, $\bar{z}(\cdot) \in Z$ is supposed to be the solution of*

$$\dot{z}(t) = \tilde{g}\Big(t, \bar{\theta}(t), z(t), \bar{u}(t)\Big), \quad t_0 \le t \le t_f, \tag{3.3a}$$

$$z(t_0) = \bar{z}_0 \tag{3.3b}$$

in the integral sense (3.2b). Then there is an open neighborhood of $\bar{\xi}$*, denoted by* $V^0(\bar{\xi})$*, such that for each open connected neighborhood* $V(\bar{\xi})$ *of* $\bar{\xi}$ *contained in* $V^0(\bar{\xi})$ *there exists a unique continuous mapping* $S : V(\bar{\xi}) \to Z$ *such that (a)* $S(\bar{\xi}) = \bar{z}(\cdot)$*; (b)* $\tau\Big(\xi, S(\xi)\Big) = 0$ *for each* $\xi \in V(\bar{\xi})$*, i.e.,* $S(\xi) = S(\xi)(t)$, $t_0 \le t \le t_f$*, is the solution of*

$$z(t) = z_0 + \int_{t_0}^{t} \tilde{g}\Big(s, \theta(s), z(s), u(s)\Big)ds, \quad t_0 \le t \le t_f, \tag{3.3c}$$

where $\xi = \Big(z_0, \theta(\cdot), u(\cdot)\Big)$*; (c) S is continuously differentiable on* $V(\bar{\xi})$*, and it holds*

$$D_u S(\xi) = -\Big(D_z \tau\big(\xi, S(\xi)\big)\Big)^{-1} D_u \tau\Big(\xi, S(\xi)\Big), \quad \xi \in V(\bar{\xi}). \tag{3.3d}$$

An immediate consequence is given next:

Corollary 3.1 *The directional derivative* $\zeta(\cdot) = \zeta_{u,h}(\cdot) = D_u S(\xi)h(\cdot)(\in Z)$, $h(\cdot) \in U$*, satisfies the integral equation*

$$\begin{aligned} \zeta(t) \; &- \int_{t_0}^{t} D_z \tilde{g}\Big(s, \theta(s), S(\xi)(s), u(s)\Big)\zeta(s)ds \\ &= \int_{t_0}^{t} D_u \tilde{g}\Big(s, \theta(s), S(\xi)(s), u(s)\Big)h(s)ds, \end{aligned} \tag{3.3e}$$

where $t_0 \le t \le t_f$ *and* $\xi = \Big(z_0, \theta(\cdot), u(\cdot)\Big)$.

Remark 3.1 Taking the time derivative of equation (3.3e) shows that this integral equation is equivalent to the so-called *perturbation equation*, see, e.g., [18].

For an arbitrary $h(\cdot) \in U$ the mappings

$$(t, \xi) \to S(\xi)(t), \; (t, \xi) \to \Big(D_u S(\xi)h(\cdot)\Big)(t), (t, \xi) \in [t_0, t_f] \times V(\bar{\xi}) \tag{3.3f}$$

are continuous and therefore also measurable.

The existence of a unique solution $\bar{z} = \bar{z}(t)$, $t_0 \le t \le t_f$, of the reference differential equation (3.3a), (3.3b) can be guaranteed as follows, where *solution* is interpreted in the integral sense, i.e., $\bar{z} = \bar{z}(t)$, $t_0 \le t \le t_f$, is absolutely continuous,

satisfies equation (3.3a) almost everywhere in the time interval $[t_0, t_f]$ and the initial condition (3.3b), cf. [7, 40].

Lemma 3.2 *Consider an arbitrary input vector* $\bar{\xi} = \Big(\bar{z}_0, \bar{\theta}(\cdot), \bar{u}(\cdot)\Big) \in \Xi$, *and define, see* (3.3a), (3.3b), *the function* $\tilde{g}_{\bar{\theta}(\cdot),\bar{u}(\cdot)} = \tilde{g}_{\bar{\theta}(\cdot),\bar{u}(\cdot)}\big(t, z\big) := \tilde{g}\Big(t, \bar{\theta}(t), z, \bar{u}(t)\Big)$. *Suppose that*

(i) $z \to \tilde{g}_{\bar{\theta}(\cdot),\bar{u}(\cdot)}\big(t, z\big)$ *is continuous for each time* $t \in [t_0, t_f]$,
(ii) $t \to \tilde{g}_{\bar{\theta}(\cdot),\bar{u}(\cdot)}\big(t, z\big)$ *is measurable for each vector* z,
(iii) generalized Lipschitz condition: For each closed sphere $K \subset \mathbb{R}^n$ *there exists a nonnegative, integrable function* $L_S(\cdot)$ *on* $[t_0, t_f]$ *such that*
(iv) $\|\tilde{g}_{\bar{\theta}(\cdot),\bar{u}(\cdot)}\big(t, 0\big)\| \le L_K(t)$, *and* $\|\tilde{g}_{\bar{\theta}(\cdot),\bar{u}(\cdot)}\big(t, z\big) - \tilde{g}_{\bar{\theta}(\cdot),\bar{u}(\cdot)}\big(t, w\big)\| \le L_K(t)$ $\|z - w\|$ *on* $[t_0, t_f] \times K$.

Then, the initial value problem (3.3a), (3.3b) *has a unique solution* $\bar{z} = \bar{z}(t; \bar{\xi})$.

Proof Proofs can be found in [7, 40]. □

We observe that the controlled stochastic process $z = z(t, \omega)$ defined by the plant differential equation (3.1a), (3.1b) may be a non Markovian stochastic process, see [3, 17]. Moreover, note that the random input function $\theta = \theta(t, \omega)$ is not just an additive noise term, but may describe also a disturbance which is part of the nonlinear dynamics of the plant, random varying model parameters such as material, load or cost parameters, etc.

3.1.1.2 Stochastic Differential Equations

In some applications [3], instead of the system (3.1a), (3.1b) of ordinary differential equations with (time-varying) random parameters, a so-called stochastic differential equation [33] is taken into account:

$$dz(t, \omega) = \tilde{g}\Big(t, z(t, \omega), u(t)\Big)dt + \tilde{h}\Big(t, z(t, \omega), u(t)\Big)d\theta(t, \omega). \tag{3.4a}$$

Here, the "noise" term $\theta = \theta(t, \omega)$ is a certain stochastic process, as, e.g., the Brownian motion, having continuous paths, and $\tilde{g} = \tilde{g}(t, z, u)$, $\tilde{h} = \tilde{h}(t, z, u)$ are given, sufficiently smooth vector/matrix functions.

Corresponding to the integral equation (3.2a), (3.2b), the above stochastic differential equation is replaced by the stochastic integral equation

$$z(t, \omega) = z_0(\omega) + \int_{t_0}^{t} \tilde{g}\Big(s, z(s, \omega), u(s)\Big)ds + \int_{t_0}^{t} \tilde{h}\Big(s, z(s, \omega), u(s)\Big)d\theta(s, \omega). \tag{3.4b}$$

The meaning of this equation depends essentially on the definition (interpretation) of the "stochastic integral"

$$I\Big(\xi(\cdot,\cdot), z(\cdot,\cdot)\Big)(t) := \int_{t_0}^{t} \tilde{h}\Big(s, z(s,\omega), u(s)\Big) d\theta(s,\omega). \tag{3.4c}$$

Note, that in case of closed-loop and open-loop feedback controls, see the next Sect. 3.2, the control function $u = u(s,\omega)$ is random. If

$$\tilde{h}(s,z,u) = \tilde{h}\Big(s, u(s)\Big) \tag{3.5a}$$

with a deterministic control $u = u(t)$ and a matrix function $\tilde{h} = \tilde{h}(s,u)$, such that $s \to \tilde{h}\Big(s, u(s)\Big)$ is differentiable, by partial integration we get, cf. [33],

$$\begin{aligned} I\Big(\xi, z(\cdot)\Big)(t) \; &= I\Big(\theta(\cdot), u(\cdot)\Big)(t) = \tilde{h}\Big(t, u(t)\Big)\theta(t) - \tilde{h}\Big(t_0, u(t_0)\Big)\theta(t_0) \\ &\quad - \int_{t_0}^{t} \theta(s) \frac{d}{ds} \tilde{h}\Big(s, u(s)\Big) ds, \end{aligned} \tag{3.5b}$$

where $\theta = \theta(s)$ denotes a sample function of the stochastic process $\theta = \theta(t,\omega)$. Hence, in this case the operator $\tau = \tau\Big(\xi, z(\cdot)\Big)$, cf. (3.2d), is defined by

$$\begin{aligned} \tau\Big(\xi, z(\cdot)\Big)(t) := z(t) - \Big(& z_0 + \int_{t_0}^{t} \tilde{g}\Big(s, z(s), u(s)\Big) ds + \tilde{h}\Big(t, u(t)\Big)\theta(t) \\ & - \tilde{h}\Big(t_0, u(t_0)\Big)\theta(t_0) - \int_{t_0}^{t} \theta(s) \frac{d}{ds} \tilde{h}\Big(s, u(s)\Big) ds \Big). \end{aligned} \tag{3.5c}$$

Obviously, for the consideration of the existence and differentiability of solutions $z(\cdot) = z(\cdot,\xi)$ of the operator equation $\tau\Big(\xi, z(\cdot)\Big) = 0$, the same procedure as in Sect. 3.1.1 may be applied.

3.1.2 *Objective Function*

The aim is to obtain optimal controls being **robust**, i.e., most insensitive with respect to stochastic variations of the model/environmental parameters and initial values of the process. Hence, incorporating stochastic parameter variations into the optimization process, for a *deterministic* control function $u = u(t)$, $t_0 \le t \le t_f$, the objective function $F = F\Big(u(\cdot)\Big)$ of the controlled process $z = z(t,\omega,u(\cdot))$ is defined, cf. [26],

by the conditional expectation of the total costs arising along the whole control process:

$$F\Big(u(\cdot)\Big) := Ef\Big(\omega, S\big(\omega, u(\cdot)\big), u(\cdot)\Big). \tag{3.6a}$$

Here, $E = E(\cdot|\mathfrak{A}_{t_0})$, denotes the conditional expectation given the information $\mathfrak{A}_{t_0}$ about the control process up to the considered starting time point t_0. Moreover, $f = f\Big(\omega, z(\cdot), u(\cdot)\Big)$ denote the stochastic total costs arising along the trajectory $z = z(t, \omega)$ and at the terminal point $z_f = z(t_f, \omega)$, cf. [3, 38]. Hence,

$$f\Big(\omega, z(\cdot), u(\cdot)\Big) := \int_{t_0}^{t_f} L\Big(t, \omega, z(t), u(t)\Big)dt + G\Big(t_f, \omega, z(t_f)\Big), \tag{3.6b}$$

$z(\cdot) \in Z, u(\cdot) \in U$. Here,

$$\begin{aligned} L &: [t_0, t_f] \times \Omega \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}, \\ G &: [t_0, t_f] \times \Omega \times \mathbb{R}^m \qquad\quad \to \mathbb{R} \end{aligned}$$

are given measurable cost functions. We suppose that $L(t, \omega, \cdot, \cdot)$ and $G(t, \omega, \cdot)$ are *convex functions* for each $(t, \omega) \in [t_0, t_f] \times \Omega$, having continuous partial derivatives $\nabla_z L(\cdot, \omega, \cdot, \cdot)$, $\nabla_u L(\cdot, \omega, \cdot, \cdot)$, $\nabla_z G(\cdot, \omega, \cdot)$. Note that in this case

$$\Big(z(\cdot), u(\cdot)\Big) \to \int_{t_0}^{t_f} L\Big(t, \omega, z(t), u(t)\Big)dt + G\Big(t_f, \omega, z(t_f)\Big) \tag{3.6c}$$

is a convex function on $Z \times U$ for each $\omega \in \Omega$. Moreover, assume that the expectation $F\Big(u(\cdot)\Big)$ exists and is finite for each admissible control $u(\cdot) \in D$.

In the case of random inputs $u = u(t, \omega), t_0 \le t \le t_f, \omega \in \Omega$, with definition (3.6b), the objective function $F = F\Big(u(\cdot, \cdot)\Big)$ reads

$$F\Big(u(\cdot, \cdot)\Big) := Ef\Big(\omega, S\Big(\omega, u(\cdot, \omega)\Big), u(\cdot, \omega)\Big). \tag{3.6d}$$

Example 3.1 (*Tracking problems*) If a trajectory $z_f = z_f(t, \omega)$, e.g., the trajectory of a moving target, known up to a certain stochastic uncertainty, must be followed or reached during the control process, then the cost function L along the trajectory can be defined by

$$L\Big(t, \omega, z(t), u\Big) := \Big\|\Gamma_z\Big(z(t) - z_f(t, \omega)\Big)\Big\|^2 + \varphi(u). \tag{3.6e}$$

In (3.6e) Γ_z is a weight matrix, and $\varphi = \varphi(u)$ denotes the control costs, as, e.g.,

$$\varphi(u) = \|\Gamma_u u\|^2 \tag{3.6f}$$

with a further weight matrix Γ_u.

If a random target $z_f = z_f(\omega)$ has to be reached at the terminal time point t_f only, then the terminal cost function G may be defined, e.g., by

$$G\Big(t_f, \omega, z(t_f)\Big) := \Big\| G_f\Big(z(t_f) - z_f(\omega)\Big)\Big\|^2 \tag{3.6g}$$

with a weight matrix G_f.

Example 3.2 (*Active structural control, control of robots*) In case of active structural control or for optimal regulator design of robots, cf. [24, 37], the total cost function f is given by defining the individual cost functions L and G as follows:

$$L(t, \omega, z, u) := z^T Q(t, \omega) z + u^T R(t, \omega) u \tag{3.6h}$$

$$G(t_f, \omega, z) := G(\omega, z). \tag{3.6i}$$

Here, $Q = Q(t, \omega)$ and $R = R(t, \omega)$, resp., are certain positive (semi)definite $m \times m, n \times n$ matrices which may depend also on (t, ω). Moreover, the terminal cost function G depends then on (ω, z). For example, in case of endpoint control, the cost function G is given by

$$G(\omega, z) = (z - z_f)^T G_f(\omega)(z - z_f) \tag{3.6j}$$

with a certain desired, possibly random terminal point $z_f = z_f(\omega)$ and a positive (semi)definite, possibly random weight matrix $G_f = G_f(\omega)$.

3.1.2.1 Optimal Control Under Stochastic Uncertainty

For finding *optimal controls being robust with respect to stochastic parameter variations* $u^\star(\cdot)$, $u^\star(\cdot, \cdot)$, resp., in this chapter we are presenting now several methods for approximation of the following minimum expected total cost problem:

$$\min F\Big(u(\cdot)\Big) \text{ s.t. } u(\cdot) \in D, \tag{3.7}$$

$$\min F\Big(u(\cdot, \cdot)\Big) \text{ s.t. } u(\cdot, \omega) \in D \quad \text{a.s. (almost sure)}, \quad u(t, \cdot)\ \mathfrak{A}_t\text{-measurable}, \tag{$\widetilde{3.7}$}$$

where $\mathfrak{A}_t \subset \mathfrak{A}$, $t \geq t_0$, denotes the σ-algebra of events $A \in \mathfrak{A}$ until time t.

Information set $\mathfrak{A}_t$ *at time* t: In many cases, as, e.g., for $PD-$ and $PID-$ controllers, see Chap. 10, the information σ-algebra $\mathfrak{A}_t$ is given by $\mathfrak{A}_t = \mathfrak{A}(y(t,\cdot))$, where $y = y(t,\omega)$ denotes the $\bar{m}-$vector function of *state-measurements or -observations* at time t. Then, an $\mathfrak{A}_t-$ measurable control $u = u(t,\omega)$ has the representation, cf. [5],

$$u(t,\omega) = \eta(t, y(t,\omega)) \tag{3.8}$$

with a measurable function $\eta(t,\cdot) : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m$.

Since parameter-insensitive optimal controls can be obtained by stochastic optimization methods incorporating random parameter variations into the optimization procedure, see [26], the aim is to determine *stochastic optimal controls*:

Definition 3.1 An optimal solution of the expected total cost minimization problem (3.7), ($\widetilde{3.7}$), resp., providing robust optimal controls, is called a **stochastic optimal control**.

Note 3.1 For controlled processes working on a time range $t_b \le t \le t_f$ with an intermediate starting time point t_b, the objective function $F = F(u(\cdot))$ is defined also by (3.6a), but with the conditional expectation operator $E = E(\cdot|\mathfrak{A}_{t_b})$, where $\mathfrak{A}_{t_b}$ denotes the information about the controlled process available up to time t_b.

Problem (3.7) is of course equivalent $\Big(E = E(\cdot|\mathfrak{A}_{t_0})\Big)$ to the *optimal control problem under stochastic uncertainty*:

$$\min E\left(\int_{t_0}^{t_f} L\Big(t,\omega,z(t),u(t)\Big)dt + G\Big(t_f,\omega,z(t_f)\Big)\Big|\mathfrak{A}_{t_0}\right) \tag{3.9a}$$

s.t.

$$\dot{z}(t) = g\Big(t,\omega,z(t),u(t)\Big), \quad t_0 \le t \le t_f, \text{ a.s.} \tag{3.9b}$$

$$z(t_0,\omega) = z_0(\omega), \text{ a.s.} \tag{3.9c}$$

$$u(\cdot) \in D, \tag{3.9d}$$

cf. [21, 22].

Remark 3.2 Similar representations can be obtained also for the second type of stochastic control problem ($\widetilde{3.7}$).

Remark 3.3 *State constraints.* In addition to the plant differential equation (dynamic equation) (3.9b), (3.9c) and the control constraints (3.9d), we may still have some stochastic state constraints

$$h_I\Big(t,\omega,z(t,\omega)\Big) \le (=)0 \text{ a.s.} \tag{3.9e}$$

as well as state constraints involving (conditional) expectations

$$Eh_{II}\Big(t,\omega,z(t,\omega)\Big) = E\Big(h_{II}\Big(t,\omega,z(t,\omega)\Big)\Big|\mathfrak{A}_{t_0}\Big) \leq (=)0. \tag{3.9f}$$

Here, $h_I = h_I(t,\omega,z)$, $h_{II} = h_{II}(t,\omega,z)$ are given vector functions of (t,ω,z). By means of (penalty) cost functions, the random condition (3.9e) can be incorporated into the objective function (3.9a). As explained in Sect. 3.8, the expectations arising in the mean value constraints (3.9f) and in the objective function (3.9a) can be computed approximatively by means of Taylor expansion with respect to the vector $\vartheta = \vartheta(\omega) := (z_0(\omega), \theta(\omega)$ of random initial values and model parameters at the conditional mean $\overline{\vartheta} = \overline{\vartheta}^{(t_0)} := E\Big(\vartheta(\omega)|\mathfrak{A}_{t_0}\Big)$. This yields then ordinary deterministic constraints for the extended deterministic trajectory (nominal state and sensitivity)

$$t \to \Big(z(t,\overline{\vartheta}),\, D_\vartheta z(t,\overline{\vartheta})\Big), \quad t \geq t_0.$$

3.2 Control Laws

Control or guidance usually refers [3, 18, 20] to direct influence on a dynamic system to achieve the desired performance. In optimal control of dynamic systems mostly the following types of *control laws* or *control policies* are considered:

(I) *Open-Loop Control (OL)*
Here, the control function $u = u(t)$ is a **deterministic** function depending only on the (a priori) information I_{t_0} about the system, the model parameters, resp., available at the starting time point t_0. Hence, for the optimal selection of optimal (OL) controls

$$u(t) = u\Big(t; t_0, I_{t_0}\Big), t \geq t_0, \tag{3.10a}$$

we get optimal control problems of type (3.7).

(II) *Closed-Loop control (CL) or Feedback Control*
In this case the control function $u = u(t)$ is a **stochastic** function

$$u = u(t,\omega) = u(t, I_t), \quad t \geq t_0 \tag{3.10b}$$

depending on time t and the total information I_t about the system available up to time t. Especially I_t may contain information about the state $z(t) = z(t,\omega)$ up to time t. Optimal (CL) or feedback controls are obtained by solving problems of type (3.7).

Remark 3.4 *Information set* $\mathfrak{A}_t$ at time t: Often the information I_t available up to time t is described by the *information set* or $\sigma-$algebra $\mathfrak{A}_t \subset \mathcal{A}$ of events A occurred

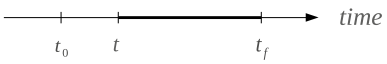


Fig. 3.1 Remaining time interval for intermediate time points t

up to time t. In the important case $\mathfrak{A}_t = \mathfrak{A}(y(t,\cdot))$, where $y = y(t,\omega)$ denotes the $\bar{m}-$vector function of *state-measurements or -observations* at time t, then an $\mathfrak{A}_t-$measurable control $u = u(t,\omega)$, see problem $(\widetilde{3.7})$, has the representation, cf. [5],

$$u(t,\omega) = \eta_t(y(t,\omega)) \tag{3.10c}$$

with a measurable function $\eta_t : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m$. Important examples of this type are the $PD-$ and $PID-$controllers, see Chap. 10.

(III) *Open-Loop Feedback (OLF) Control/Stochastic Open-Loop Feedback (SOLF) Control*
Due to their large complexity, in general, optimal feedback control laws can be determined approximatively only. A very efficient approximation procedure for optimal feedback controls, being functions of the information I_t, is the approximation by means of optimal open-loop controls. In this combination of (OL) and (CL) control, at each intermediate time point $t_b := t$, $t_0 \le t \le t_f$, given the information I_t up to time t, first the open-loop control function for the remaining time interval $t \le s \le t_f$, see Fig. 3.1, is computed, hence,

$$u_{[t,t_f]}(s) = u\Big(s; t, I_t\Big), \quad s \ge t. \tag{3.10d}$$

Then, an approximate feedback control policy, originally proposed by Dreyfus (1964), cf. [10], can be defined as follows:

Definition 3.2 The hybrid control law, defined by

$$\varphi(t, I_t) := u\Big(t; t, I_t\Big), t \ge t_0 \tag{3.10e}$$

is called *open-loop feedback (OLF) control law*.

Thus, the OL control $u_{[t,t_f]}(s)$, $s \ge t$, for the remaining time interval $[t, t_f]$ is used only at time $s = t$, see also [2, 9–11, 15, 19, 39]. Optimal (OLF) controls are obtained therefore by solving again control problems of the type (3.7) at each intermediate starting time point $t_b := t$, $t \in [t_0, t_f]$.

A major issue in optimal control is the **robustness**, cf. [12], i.e., the insensitivity of an optimal control with respect to parameter variations. In case of random parameter variations robust optimal controls can be obtained by means of stochastic optimization methods, cf. [26], incorporating the probability distribution, i.e., the random

characteristics, of the random parameter variation into the optimization process, cf. Definition 3.1.

Thus, constructing stochastic optimal open-loop feedback controls, hence, optimal open-loop feedback control laws being insensitive as far as possible with respect to random parameter variations, means that besides the optimality of the control policy also its insensitivity with respect to stochastic parameter variations should be guaranteed. Hence, in the following sections we also develop a **stochastic version** of the optimal open-loop feedback control method, cf. [25, 27–29]. A short overview on this novel stochastic optimal open-loop feedback control concept is given below:

At each intermediate time point $t_b = t \in [t_0, t_f]$, based on the given process observation I_t, e.g., the observed state $z_t = z(t)$ at $t_b = t$, a stochastic optimal open-loop control $u^* = u^*(s) = u^*(s; t, I_t), t \le s \le t_f$, is determined first on the remaining time interval $[t, t_f]$, see Fig. 3.1, by stochastic optimization methods, cf. [26] .

Having a stochastic optimal open-loop control $u^* = u^*(s; t, I_t), t \le s \le t_f$, on each remaining time interval $[t, t_f]$ with an arbitrary starting time point t, $t_0 \le t \le t_f$, a *stochastic optimal open-loop feedback (SOLF) control law* is then defined—corresponding to Definition 3.2—as follows:

Definition 3.3 The hybrid control law, defined by

$$\varphi^*(t, I_t) := u^*\Big(t; t, I_t\Big), \quad t \ge t_0. \tag{3.10f}$$

is called the *stochastic optimal open-loop feedback (SOLF) control law*.

Thus, at time $t_b = t$ just the "first" control value $u^*(t) = u^*(t; t, I_t)$ of $u^* = u^*(\cdot; t, I_t)$ is used only.

For finding stochastic optimal open-loop controls, on the remaining time intervals $t_b \le t \le t_f$ with $t_0 \le t_b \le t_f$, the stochastic Hamilton function of the control problem is introduced. Then, the class of H-minimal controls, cf. [18], can be determined in case of stochastic uncertainty by solving a finite-dimensional stochastic optimization problem for minimizing the conditional expectation of the stochastic Hamiltonian subject to the remaining deterministic control constraints at each time point t. Having a H-minimal control, the related two-point boundary value problem with random parameters will be formulated for the computation of a stochastic optimal state- and costate-trajectory. In the important case of a linear-quadratic structure of the underlying control problem, the state and costate trajectory can be **determined analytically** to a large extent. Inserting then these trajectories into the H-minimal control, stochastic optimal open-loop controls are found on an arbitrary remaining time interval. According to Definition 3.2, these controls yield then immediately a stochastic optimal open-loop feedback control law. Moreover, the obtained controls can be realized in **real-time**, which is already shown for applications in optimal control of industrial robots, cf. [35].

(III.1) *Nonlinear Model Predictive Control (NMPC)/Stochastic Nonlinear Model Predictive Control (SNMPC)*
Optimal open-loop feedback (OLF) control is the basic tool in *Nonlinear Model Predictive Control (NMPC)*. Corresponding to the approximation technique for feedback controls described above, (NMPC) is a method to solve complicated feedback control problems by means of stepwise computations of open-loop controls. Hence, in (NMPC), see [1, 13, 14, 34] optimal open-loop controls

$$u = u_{[t,t+T_p]}(s), \quad t \le s \le t + T_p, \tag{3.10g}$$

cf. (3.8), are determined first on the time interval $[t, t + T_p]$ with a certain so-called *prediction time horizon* $T_p > 0$. In sampled-data MPC, cf. [13], optimal open-loop controls $u = u_{[t_i,t_i+T_p]}$, are determined at certain sampling instants t_i, $i = 0, 1, \ldots$, using the information $\mathfrak{A}_{t_i}$ about the control process and its neighborhood up to time t_i, $i = 0, 1, \ldots$, see also [24]. The optimal open-loop control at stage "i" is applied then,

$$u = u_{[t_i,t_i+T_p]}(t), \quad t_i \le t \le t_{i+1}, \tag{3.10h}$$

until the next sampling instant t_{i+1}. This method is closely related to the *Adaptive Optimal Stochastic Trajectory Planning and Control (AOSTPC) procedure* described in [23, 24].

Corresponding to the extension of (OLF) control to (SOLF) control, (NMPC) can be extended to *Stochastic Nonlinear Model Predictive Control (SNMPC)*. For control policies of this type, robust (NMPC) with respect to stochastic variations of model parameters and initial values are determined in the following way:

- Use the a posteriori distribution $P(d\omega|\mathfrak{A}_t)$ of the basic random element $\omega \in \Omega$, given the process information $\mathfrak{A}_t$ up to time t, and
- apply stochastic optimization methods to incorporate random parameter variations into the optimal (NMPC) control design.

3.3 Convex Approximation by Inner Linearization

We observe first that (3.7), ($\widetilde{3.7}$), resp., is in general a non-convex optimization problem, cf. [20]. Since for convex (deterministic) optimization problems there is a well-established theory, we approximate the original problems (3.7), ($\widetilde{3.7}$) by a sequence of suitable convex problems. In the following we describe first a single step of this procedure. Due to the consideration in Sect. 3.2, we may concentrate here to problem (3.7) or (3.9a)–(3.9d) for deterministic controls $u(\cdot)$, as needed in the computation of optimal (OL), (OLF), (SOLF) as well as (NMP), (SNMP) controls being most important for practical problems.

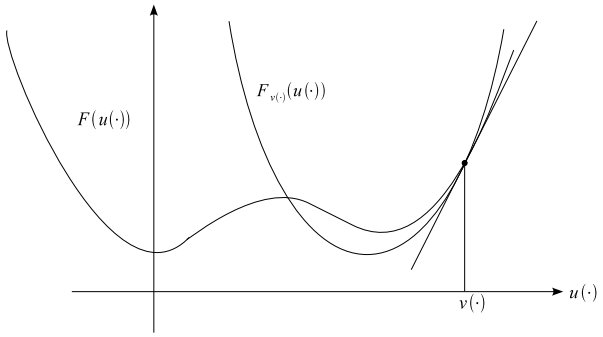


Fig. 3.2 Convex approximation

Let $v(\cdot) \in D$ be an arbitrary, but fixed admissible initial or reference control and assume, see Lemma 3.1, for the input-output map $z(\cdot) = S(\omega, u(\cdot))$:

Assumption 3.1 $S(\omega, \cdot)$ is F-differentiable at $v(\cdot)$ for each $\omega \in \Omega$.

Denote the F-derivative of $S(\omega, \cdot)$ at $v(\cdot)$ by $DS\Big(\omega, v(\cdot)\Big)$ and replace the cost function $F = F\Big(u(\cdot)\Big)$ with

$$F_{v(\cdot)}\Big(u(\cdot)\Big) := Ef\Big(\omega, S\big(\omega, v(\cdot)\big) + DS\big(\omega, v(\cdot)\big)\big(u(\cdot) - v(\cdot)\big), u(\cdot)\Big) \tag{3.11}$$

where $u(\cdot) \in U$, cf. Fig. 3.2. Assume that $F_{v(\cdot)}\Big(u(\cdot)\Big) \in \mathbb{R}$ for all pairs $\Big(u(\cdot), v(\cdot)\Big) \in D \times D$.

Then, replace the optimization problem (3.7), see [21], by

$$\min F_{v(\cdot)}\Big(u(\cdot)\Big) \text{ s.t. } u(\cdot) \in D. \tag{3.7}_{v(\cdot)}$$

Lemma 3.3 $(3.7)_{v(\cdot)}$ *is a convex optimization problem.*

Proof According to Sect. 3.1, function (3.6c) is convex. The assertion follows now from the linearity of the F-differential of $S(\omega, \cdot)$. □

Remark 3.5 Note that the approximate $F_{v(\cdot)}$ of F is obtained from (3.6a), (3.6b) by means of linearization of the input-output map $S = S\Big(\omega, u(\cdot)\Big)$ with respect to the control $u(\cdot)$ at $v(\cdot)$, hence, by **inner linearization** of the control problem with respect to the control $u(\cdot)$ at $v(\cdot)$.

Remark 3.6 (*Linear input-output map*) In case that $S = S(\omega, u(\cdot)) := S(\omega)u(\cdot)$ is linear with respect to the control $u(\cdot)$, then $DS(\omega, v(\cdot) = S(\omega)$ and we have $F_{v(\cdot)}\Big(u(\cdot)\Big) = F(u(\cdot))$. In this case the problems (3.7) and $(3.7)_{v(\cdot)}$ coincide for each input vector $v(\cdot)$.

For a real-valued convex function $\phi : X \to \mathbb{R}$ on a linear space X the directional derivative $\phi'_+(x; y)$ exists, see, e.g., [16], at each point $x \in X$ and in each direction $y \in X$. According to Lemma 3.3 the objective function $F_{v(\cdot)}$ of the approximate problem $(3.7)_{v(\cdot)}$ is convex. Using the theorem of the *monotone convergence*, [5], for all $u(\cdot), v(\cdot) \in D$ and $h(\cdot) \in U$ the directional derivative of $F_{v(\cdot)}$ is given, see [22], Satz 1.4, by

$$F'_{v(\cdot)+}\Big(u(\cdot); h(\cdot)\Big) = Ef'_+\Big(\omega, S\Big(\omega, v(\cdot)\Big) + DS\Big(\omega, v(\cdot)\Big)\Big(u(\cdot) - v(\cdot)\Big), u(\cdot); DS\Big(\omega, v(\cdot)\Big)h(\cdot), h(\cdot)\Big). \tag{3.12a}$$

In the special case $u(\cdot) = v(\cdot)$ we get

$$F'_{v(\cdot)+}\Big(v(\cdot); h(\cdot)\Big) = Ef'_+\Big(\omega, S\Big(\omega, v(\cdot)\Big), v(\cdot); DS\Big(\omega, v(\cdot)\Big)h(\cdot), h(\cdot)\Big). \tag{3.12b}$$

A solution $\bar{u}(\cdot) \in D$ of the convex problem $(3.7)_{v(\cdot)}$ is then characterized cf. [30], by

$$F'_{v(\cdot)+}\Big(\bar{u}(\cdot); u(\cdot) - \bar{u}(\cdot)\Big) \geq 0 \quad \text{for all } u(\cdot) \in D. \tag{3.13}$$

Definition 3.4 For each $v(\cdot) \in D$, let $M\Big(v(\cdot)\Big)$ be the set of solutions of problem $(2.7)_{v(\cdot)}$, i.e.,

$$M\Big(v(\cdot)\Big) := \Big\{u^0(\cdot) \in D : F'_{v(\cdot)+}\Big(u^0(\cdot); u(\cdot) - u^0(\cdot)\Big) \geq 0, u(\cdot) \in D\Big\}.$$

Note 3.2 If the input-output operator $S = S(\omega, u(\cdot)) := S(\omega)u(\cdot)$ is linear, then $M\Big(v(\cdot)\Big) = M$ for each input $v(\cdot)$, where M denotes the set of solutions of problem (3.7).

In the following we suppose that optimal solutions of $(2.7)_{v(\cdot)}$ exist for each $v(\cdot)$.

Assumption 3.2 $M\Big(v(\cdot)\Big) \neq \emptyset$ for each $v(\cdot) \in D$.

A first relation between our original problem (3.7) and the family of its approximates $(3.7)_{v(\cdot)}$, $v(\cdot) \in D$, is shown in the following:

Theorem 3.1 *Suppose that the directional derivative* $F'_+ = F'_+\Big(v(\cdot); h(\cdot)\Big)$ *exists and*

$$F'_+\Big(v(\cdot); h(\cdot)\Big) = F'_{v(\cdot)+}\Big(v(\cdot); h(\cdot)\Big) \tag{3.14}$$

for each $v(\cdot) \in D$ *and* $h(\cdot) \in D - D$. *Then:*

(I) *If $\bar{u}(\cdot)$ is an optimal control, then $\bar{u}(\cdot) \in M\Big(\bar{u}(\cdot)\Big)$, i.e., $\bar{u}(\cdot)$ is a solution of $(3.7)_{\bar{u}(\cdot)}$.*

(II) *If (3.7) is convex, then $\bar{u}(\cdot)$ is an optimal control if and only if $\bar{u}(\cdot) \in M\Big(\bar{u}(\cdot)\Big)$.*

Proof Because of the convexity of the approximate control problem $(3.7)_{v(\cdot)}$, the condition $v(\cdot) \in M\Big(v(\cdot)\Big)$ holds, cf. (3.13), if and only if $F'_{v(\cdot)+}\Big(v(\cdot); u(\cdot) - v(\cdot)\Big) \geq 0$ for all $u(\cdot) \in D$. Because of (3.14), this is equivalent with $F'_{+}\Big(v(\cdot); u(\cdot) - v(\cdot)\Big) \geq 0$ for all $u(\cdot) \in D$. However, since the admissible control domain D is convex, for an optimal solution $v(\cdot) := \bar{u}(\cdot)$ of (3.7) this condition is necessary, and necessary as also sufficient in case that (3.7) is convex. □

Assuming that $f = f(\omega, \cdot, \cdot)$ is F-differentiable for each ω, by means of (3.12b) and the chain rule we have

$$\begin{aligned} F'_{v(\cdot)+}\Big(v(\cdot); h(\cdot)\Big) &= E f'_{+}\Big(\omega, S\Big(\omega, v(\cdot)\Big), v(\cdot); DS\Big(\omega, v(\cdot)\Big)h(\cdot), h(\cdot)\Big) \\ &= E \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \Big(f\Big(\omega, S\big(\omega, v(\cdot) + \varepsilon h(\cdot)\big), v(\cdot) + \varepsilon h(\cdot)\Big) \\ &\quad - f\Big(\omega, S\big(\omega, v(\cdot)\big), v(\cdot)\Big)\Big). \end{aligned} \tag{3.15}$$

Note 3.3 Because of the properties of the operator $S = S(\omega, u(\cdot))$, the above equations holds also for arbitrary convex functions f such that the expectations under consideration exist, see [22].

Due to the definition (3.6a) of the objective function $F = F(u(\cdot))$ for condition (3.14) the following criterion holds.

Lemma 3.4

(a) *Condition (3.14) in Theorem 3.1 holds if and only if the expectation operator "E" and the limit process "$\lim_{\varepsilon \downarrow 0}$" in (3.15) may be interchanged.*

(b) *This interchangeability holds, e.g., if $\sup\Big\{\big\|DS\big(\omega, v(\cdot) + \varepsilon h(\cdot)\big)\big\| : 0 \leq \varepsilon \leq 1\Big\}$ is bounded with probability one, and the convex function $f(\omega, \cdot, \cdot)$ satisfies a Lipschitz condition*

$$\Big|f\Big(\omega, z(\cdot), u(\cdot)\Big) - f\Big(\omega, \bar{z}(\cdot), \bar{u}(\cdot)\Big)\Big| \leq \gamma(\omega) \Big\|\Big(z(\cdot), u(\cdot)\Big) - \Big(\bar{z}(\cdot), \bar{u}(\cdot)\Big)\Big\|_{Z \times U}$$

on a set $Q \subset Z \times U$ containing all vectors $\Big(S\big(\omega, v(\cdot) + \varepsilon h(\cdot)\big), v(\cdot) + \varepsilon h(\cdot)\Big)$, $0 \leq \varepsilon \leq 1$, where $E\gamma(\omega) < +\infty$, and $\|\cdot\|_{Z \times U}$ denotes the norm on $Z \times U$.

Proof The first assertion (a) follows immediately from (3.15) and the definition of the objective function F. Assertion (b) can be obtained by means of the generalized mean value theorem for vector functions and Lebesgue's bounded convergence theorem. □

Remark 3.7 Further conditions are given in [22].

A second relation between our original problem (3.7) and the family of its approximates $(3.7)_{v(\cdot)}$, $v(\cdot) \in D$, is shown next.

Lemma 3.5

(a) If $\bar{u}(\cdot) \notin M\left(\bar{u}(\cdot)\right)$ *for a control* $\bar{u}(\cdot) \in D$, *then*

$$F_{\bar{u}(\cdot)}\left(u(\cdot)\right) < F_{\bar{u}(\cdot)}\left(\bar{u}(\cdot)\right) = F\left(\bar{u}(\cdot)\right) \text{ for each } u(\cdot) \in M\left(\bar{u}(\cdot)\right)$$

(b) Let the controls $u(\cdot), v(\cdot) \in D$ *be related such that*

$$F_{u(\cdot)}\left(v(\cdot)\right) < F_{u(\cdot)}\left(u(\cdot)\right).$$

If (3.14) holds for the pair $\left(u(\cdot), h(\cdot)\right)$, $h(\cdot) = v(\cdot) - u(\cdot)$, *then* $h(\cdot)$ *is an admissible direction of decrease for* F *at* $u(\cdot)$, *i.e., we have* $F\left(u(\cdot) + \varepsilon h(\cdot)\right) < F\left(u(\cdot)\right)$ *and* $u(\cdot) + \varepsilon h(\cdot) \in D$ *on a suitable interval* $0 < \varepsilon < \bar{\varepsilon}$.

Proof According to Definition 3.4, for $u(\cdot) \in M\left(\bar{u}(\cdot)\right)$ we have $F_{\bar{u}(\cdot)}(u(\cdot)) \leq F_{\bar{u}(\cdot)}(v(\cdot))$ for all $v(\cdot) \in D$ and therefore also $F_{\bar{u}(\cdot)}(u(\cdot)) \leq F_{\bar{u}(\cdot)}(\bar{u}(\cdot))$. a) In case $F_{\bar{u}(\cdot)}(u(\cdot)) = F_{\bar{u}(\cdot)}(\bar{u}(\cdot))$ we get $F_{\bar{u}(\cdot)}(\bar{u}(\cdot)) \leq F_{\bar{u}(\cdot)}(v(\cdot))$ for all $v(\cdot) \in D$, hence, $\bar{u}(\cdot) \in M\left(\bar{u}(\cdot)\right)$. Since this is in contradiction to the assumption, it follows $F_{\bar{u}(\cdot)}(u(\cdot)) < F_{\bar{u}(\cdot)}(\bar{u}(\cdot))$. b) If controls $u(\cdot), v(\cdot) \in D$ are related $F_{u(\cdot)}\left(v(\cdot)\right) < F_{u(\cdot)}\left(u(\cdot)\right)$, then due to the convexity of $F_{u(\cdot)}$ we have $F'_{u(\cdot)+}\left(u(\cdot); v(\cdot) - u(\cdot)\right) \leq F_{u(\cdot}(v(\cdot)) - F_{u(\cdot)}(u(\cdot)) < 0$. With (3.14) we then get $F'_{+}\left(u(\cdot); v(\cdot) - u(\cdot)\right) = F'_{u(\cdot)+}\left(u(\cdot); v(\cdot) - u(\cdot)\right) < 0$. This yields now that $h(\cdot) := v(\cdot) - u(\cdot)$ is a feasible descent direction for F at $u(\cdot)$. □

If $\bar{u}(\cdot) \in M\left(\bar{u}(\cdot)\right)$, then the convex approximate $F_{\bar{u}}$ of F at $\bar{u}$ cannot be decreased further on D. Thus, the above results suggest the following definition:

Definition 3.5 A control $\bar{u}(\cdot) \in D$ such that $\bar{u}(\cdot) \in M\left(\bar{u}(\cdot)\right)$ is called a **stationary control** of the optimal control problem (3.7).

Under the rather weak assumptions in Theorem 3.1 an optimal control is also stationary, and in the case of a convex problem (3.7) the two concepts coincide. Hence, stationary controls are candidates for optimal controls. As an appropriate substitute/approximate for an optimal control we may determine therefore stationary controls. For this purpose algorithms of the following *conditional gradient-type* may be applied:

Algorithm 3.1 (I) Choose $u^1 \in D$, put $j = 1$

(II) If $u^j(\cdot) \in M\Big(u^j(\cdot)\Big)$, then $u^j(\cdot)$ is stationary and the algorithm stops; otherwise find a control $v^j(\cdot) \in M\Big(u^j(\cdot)\Big)$

(III) Set $u^{j+1}(\cdot) = v^j(\cdot)$ and go to II), putting $j \to j+1$.

Algorithm 3.2 (I) Choose $u^1(\cdot) \in D$, put $j = 1$

(II) If $u^j(\cdot) \in M\Big(u^j(\cdot)\Big)$, then $u^j(\cdot)$ is stationary and the algorithm stops; otherwise find a $v^{(j)}(\cdot) \in M\Big(u^j(\cdot)\Big)$, define $h^j(\cdot) := v^j(\cdot) - u^j(\cdot)$

(III) Calculate $\bar{u}(\cdot) \in m\Big(u^j(\cdot), h^j(\cdot)\Big)$, set $u^{j+1}(\cdot) := \bar{u}(\cdot)$ and go to II), putting $j \to j+1$.

Here, based on line search, $m\Big(u(\cdot), h(\cdot)\Big)$ is defined by

$$m\Big(u(\cdot), h(\cdot)\Big) = \Big\{u(\cdot) + \varepsilon^* h(\cdot) : F\Big(u(\cdot) + \varepsilon^* h(\cdot)\Big) = \min_{0 \le \varepsilon \le 1} F\Big(u(\cdot) + \varepsilon h(\cdot)\Big) \text{ for } \varepsilon^* \in [0,1]\Big\}, u(\cdot), h(\cdot) \in U.$$

Concerning *Algorithm* 3.1 we have the following result.

Theorem 3.2 *Let the set valued mapping $u(\cdot) \to M\Big(u(\cdot)\Big)$ be closed at each $\bar{u}(\cdot) \in D\Big($,i.e., the relations $u^j(\cdot) \to \bar{u}(\cdot)$, $v^j(\cdot) \in M\Big(u^j(\cdot)\Big)$, $j = 1, 2 \ldots$, and $v^j(\cdot) \to \bar{v}(\cdot)$ imply that also $\bar{v}(\cdot) \in M\Big(\bar{u}(\cdot)\Big)\Big)$.*

If a sequence $u^1(\cdot), u^2(\cdot), \ldots$ of controls generated by Algorithm 3.1 converges to an element $\bar{u}(\cdot) \in D$, then $\bar{u}(\cdot)$ is a stationary control.

A sufficient condition for the closedness of the algorithmic map $u(\cdot) \to M\Big(u(\cdot)\Big)$ is given next:

Lemma 3.6 *Let D be a closed set of admissible controls, and let*

$$\Big(u(\cdot), v(\cdot)\Big) \to F'_{u(\cdot)+}\Big(v(\cdot); w(\cdot) - v(\cdot)\Big)$$

be continuous on $D \times D$ for each $w(\cdot) \in D$. Then $u(\cdot) \to M\Big(u(\cdot)\Big)$ is closed at each element of D.

While the convergence assumption for a sequence $u^1(\cdot), u^2(\cdot), \ldots$ generated by *Algorithm* 3.1 is rather strong, only the existence of accumulation points of $u^j(\cdot)$, $j = 1, 2, \ldots$, has to be required in *Algorithm* 3.2.

3.4 Computation of Directional Derivatives

Suppose here again that $U := L_\infty^n([t_0, t_f], \mathcal{B}^1, \lambda^1)$ is the Banach space of all essentially bounded measurable functions $u(\cdot) : [t_0, t_f] \to \mathbb{R}^n$, normed by the essential supremum norm. According to Definitions 3.4, 3.5 of a stationary control and characterization (3.13) of an optimal solution of $(3.7)_{v(\cdot)}$, we first have to determined the directional derivative $F'_{v(\cdot)+}$. Based on the justification in Sect. 3.1, we assume again that the solution $z(t, \omega) = S\Big(\omega, u(\cdot)\Big)(t)$ of (3.2b) is measurable in $(t, \omega) \in [t_0, t_f] \times \Omega$ for each $u(\cdot) \in D$, and $u(\cdot) \to S\Big(\omega, u(\cdot)\Big)$ is continuously differentiable on D for each $\omega \in \Omega$. Furthermore, we suppose that the F-differential $\zeta(t) = \zeta(t, \omega) = \Big(D_u S\Big(\omega, u(\cdot)\Big)h(\cdot)\Big)(t)$, $h(\cdot) \in U$, is measurable and essentially bounded in (t, ω), and is given according to (3.3a)–(3.3f) by the linear integral equation

$$\zeta(t) - \int_{t_0}^{t} A\Big(t, \omega, u(\cdot)\Big)\zeta(s)ds = \int_{t_0}^{t} B\Big(t, \omega, u(\cdot)\Big)h(s)ds, \qquad (3.16a)$$

$$t_0 \le t \le t_f,$$

with the Jacobians

$$A\Big(t, \omega, u(\cdot)\Big) := D_z g\Big(t, \omega, z_u(t, \omega), u(t)\Big) \qquad (3.16b)$$

$$B\Big(t, \omega, u(\cdot)\Big) := D_u g\Big(t, \omega, z_u(t, \omega), u(t)\Big) \qquad (3.16c)$$

and $z_u = z_u(t, \omega)$ defined, cf. (3.3f), by

$$z_u(t, \omega) := S\Big(\omega, u(\cdot)\Big)(t), \quad \Big(t, \omega, u(\cdot)\Big) \in [t_0, t_f] \times \Omega \times U. \qquad (3.16d)$$

Here, the random element "ω" is also used, cf. Sect. 3.1.1, to denote the realization of the random inputs $z_0 = z_0(\omega)$, $\theta(\cdot) = \theta(\cdot, \omega)$.

$$\omega := \begin{pmatrix} z_0 \\ \theta(\cdot) \end{pmatrix}$$

Remark 3.8 Due to the measurability of the functions $z_u = z_u(t, \omega)$ and $u = u(t)$ on $[t_0, t_f] \times \Omega$, $[t_0, t_f]$, resp., and the assumptions on the function g and its Jacobians $D_z g$, $D_u g$, see Sect. 3.1, also the matrix-valued functions $(t, \omega) \to A(t, \omega, u(\cdot))$, $(t, \omega) \to B(t, \omega, u(\cdot))$ are measurable and essentially bounded on $[t_0, t_f] \times \Omega$. Equation (3.16a) is again a vectorial Volterra integral equation, and the existence

of a unique measurable solution $\zeta(t) = \zeta(t, \omega)$ can be shown as for the Volterra integral equation (3.2g), 3.2h).

The differential form of (3.16a) is then the *linear perturbation equation*

$$\dot{\zeta}(t) = A\Big(t, \omega, u(\cdot)\Big)\zeta(t) + B\Big(t, \omega, u(\cdot)\Big)h(t), t_0 \le t \le t_f, \omega \in \Omega \tag{3.16e}$$

$$\zeta(t_0) = 0. \tag{3.16f}$$

The solution $\zeta = \zeta(t, \omega)$ of (3.16a), (3.16e), (3.16f), resp., is also denoted, cf. (3.3f), by

$$\zeta(t, \omega) = \zeta_{u,h}(t, \omega) := \Big(D_u S\Big(\omega, u(\cdot)\Big)h(\cdot)\Big)(t), h(\cdot) \in U. \tag{3.16g}$$

This means that the approximate $(3.7)_{v(\cdot)}$ of (3.7) has the following explicit form:

$$\min E\Big(\int_{t_0}^{t_f} L\Big(t, \omega, z_v(t, \omega) + \zeta(t, \omega), u(t)\Big)dt + G\Big(t_f, \omega, z_v(t_f, \omega) + \zeta(t_f, \omega)\Big)\Big) \tag{3.17a}$$

s.t.

$$\dot{\zeta}(t, \omega) = A\Big(t, \omega, v(\cdot)\Big)\zeta(t, \omega) + B\Big(t, \omega, v(\cdot)\Big)\Big(u(t) - v(t)\Big) \text{a.s.} \tag{3.17b}$$

$$\zeta(t_0, \omega) = 0 \text{ a.s.} \tag{3.17c}$$

$$u(\cdot) \in D. \tag{3.17d}$$

With the convexity assumptions in Sect. 3.1.2, Lemma 3.3 yields that (3.17a)–(3.17d) is a convex stochastic control problem, with a *linear* plant differential equation.

For the subsequent analysis of the stochastic control problem we need now a representation of the directional derivative $F'_{v(\cdot)+}\Big(u(\cdot); h(\cdot)\Big)$ by a scalar product

$$F'_{v(\cdot)+}\Big(u(\cdot); h(\cdot)\Big) = \int_{t_0}^{t_f} q(t)^T h(t)dt$$

with a certain deterministic vector function $q = q(t)$. From representation (3.12a) of the directional derivative $F'_{v(\cdot)+}$ of the convex approximate $F_{v(\cdot)}$ of F, definition (3.6b) of $f = f\Big(\omega, z(\cdot), u(\cdot)\Big)$ by an integral over $[t_0, t_f]$ and [22], Satz 1.4, with (3.16g) we obtain

$$F'_{v(\cdot)+}\Big(u(\cdot);h(\cdot)\Big) = E\Bigg(\int_{t_0}^{t_f}\Big(\nabla_z L\Big(t,\omega,z_v(t,\omega)+\zeta_{v,u-v}(t,\omega),u(t)\Big)^T\zeta_{v,h}(t,\omega) + \nabla_u L\Big(t,\omega,z_v(t,\omega)+\zeta_{v,u-v}(t,\omega),u(t)\Big)^T h(t)\Big) + \nabla_z G\Big(t_f,\omega,z_v(t_f,\omega)+\zeta_{v,u-v}(t_f,\omega)\Big)^T\zeta_{v,h}(t_f,\omega)\Bigg). \quad (3.18)$$

Defining the gradients

$$a\Big(t,\omega,v(\cdot),u(\cdot)\Big) := \nabla_z L\Big(t,\omega,z_v(t,\omega)+\zeta_{v,u-v}(t,\omega),u(t)\Big) \quad (3.19a)$$

$$b\Big(t,\omega,v(\cdot),u(\cdot)\Big) := \nabla_u L\Big(t,\omega,z_v(t,\omega)+\zeta_{v,u-v}(t,\omega),u(t)\Big) \quad (3.19b)$$

$$c\Big(t_f,\omega,v(\cdot),u(\cdot)\Big) := \nabla_z G\Big(t_f,\omega,z_v(t_f,\omega)+\zeta_{v,u-v}(t_f,\omega)\Big), \quad (3.19c)$$

measurable with respect to (t,ω), the directional derivative $F'_{v(\cdot)+}$ can be represented by

$$F'_{v(\cdot)+}\Big(u(\cdot);h(\cdot)\Big) = E\Bigg(\int_{t_0}^{t_f}\Big(a\Big(t,\omega,v(\cdot),u(\cdot)\Big)^T\zeta_{v,h}(t,\omega) + b\Big(t,\omega,v(\cdot),u(\cdot)\Big)^T h(t)\Big)dt + c\Big(t_f,\omega,v(\cdot),u(\cdot)\Big)^T\zeta_{v,h}(t_f,\omega)\Bigg). \quad (3.20a)$$

According to (3.16a), for $\zeta_{v,h} = \zeta_{v,h}(t,\omega)$ we have

$$\zeta_{v,h}(t_f,\omega) = \int_{t_0}^{t_f}\Big(A\Big(t,\omega,v(\cdot)\Big)\zeta_{v,h}(t,\omega) + B\Big(t,\omega,v(\cdot)\Big)h(t)\Big)dt. \quad (3.20b)$$

Putting (3.20b) into (3.20a), we find

$$F'_{v(\cdot)+}\Big(u(\cdot);h(\cdot)\Big) = E\Bigg(\int_{t_0}^{t_f}\tilde{a}\Big(t,\omega,v(\cdot),u(\cdot)\Big)^T\zeta_{v,h}(t,\omega)dt + \int_{t_0}^{t_f}\tilde{b}\Big(t,\omega,v(\cdot),u(\cdot)\Big)^T h(t)dt\Bigg), \quad (3.20c)$$

where

$$\tilde{a}\Big(t,\omega,v(\cdot),u(\cdot)\Big) := a\Big(t,\omega,v(\cdot),u(\cdot)\Big) + A\Big(t,\omega,v(\cdot)\Big)^T c\Big(t_f,\omega,v(\cdot),u(\cdot)\Big) \quad (3.20d)$$

$$\tilde{b}\Big(t,\omega,v(\cdot),u(\cdot)\Big) := b\Big(t,\omega,v(\cdot),u(\cdot)\Big) + B\Big(t,\omega,v(\cdot)\Big)^T c\Big(t_f,\omega,v(\cdot),u(\cdot)\Big). \quad (3.20e)$$

Remark 3.9 According to Remark 3.8 also the functions $(t,\omega) \to \tilde{a}\Big(t,\omega,v(\cdot),u(\cdot)\Big)$, $(t,\omega) \to \tilde{b}\Big(t,\omega,v(\cdot),u(\cdot)\Big)$ are measurable on $[t_0,t_f] \times \Omega$.

In order to transform the first integral in (3.20c) into the form of the second integral in (3.20c), we introduce the m-vector function

$$\lambda = \lambda_{v,u}(t,\omega)$$

defined by the following integral equation depending on the random parameter ω:

$$\lambda(t) - A\Big(t,\omega,v(\cdot)\Big)^T \int_t^{t_f} \lambda(s)ds = \tilde{a}\Big(t,\omega,v(\cdot),u(\cdot)\Big). \quad (3.21)$$

Under the present assumptions, this *Volterra integral equation* has [22] a unique measurable solution $(t,\omega) \to \lambda_{v,u}(t,\omega)$, see also Remark 3.8. By means of (3.21) we obtain

$$\begin{aligned}
&\int_{t_0}^{t_f} \tilde{a}\Big(t,\omega,v(\cdot),u(\cdot)\Big)^T \zeta_{v,h}(t,\omega)dt \\
&= \int_{t_0}^{t_f} \left(\lambda(t) - A\Big(t,\omega,v(\cdot)\Big)^T \int_t^{t_f} \lambda(s)ds\right)^T \zeta_{v,h}(t,\omega)dt \\
&= \int_{t_0}^{t_f} \lambda(t)^T \zeta_{v,h}(t,\omega)dt \\
&\quad - \int_{t_0}^{t_f} dt \int_{t_0}^{t_f} ds\, J(s,t)\lambda(s)^T A\Big(t,\omega,v(\cdot)\Big)\zeta_{v,h}(t,\omega) \\
&= \int_{t_0}^{t_f} \lambda(s)^T \zeta_{v,h}(s,\omega)ds - \int_{t_0}^{t_f} ds \int_{t_0}^{t_f} dt\, J(s,t)\lambda(s)^T A\Big(t,\omega,v(\cdot)\Big)\zeta_{v,h}(t,\omega) \\
&= \int_{t_0}^{t_f} \lambda(s)^T \left(\zeta_{v,h}(s,\omega) - \int_{t_0}^{s} A\Big(t,\omega,v(\cdot)\Big)\zeta_{v,h}(t,\omega)dt\right)ds, \qquad (3.22a)
\end{aligned}$$

where $J = J(s,t)$ is defined by

$$J(s,t) := \begin{cases} 0, & t_0 \le s \le t \\ 1, & t < s \le t_f. \end{cases}$$

Using now again the perturbation equation (3.16a), (3.16b), from (3.22a) we get

$$\begin{aligned} &\int_{t_0}^{t_f} \tilde{a}\Big(t,\omega,v(\cdot),u(\cdot)\Big)^T \zeta_{v,h}(t,\omega)dt = \int_{t_0}^{t_f} \lambda(s)^T \Big(\int_{t_0}^{s} B\Big(t,\omega,v(\cdot)\Big)h(t)dt\Big)ds \\ &= \int_{t_0}^{t_f} ds\lambda(s)^T \int_{t_0}^{t_f} dt J(s,t) B\Big(t,\omega,v(\cdot)\Big)h(t) = \int_{t_0}^{t_f} dt \int_{t_0}^{t_f} ds J(s,t)\lambda(s)^T B\Big(t,\omega,v(\cdot)\Big)h(t) \\ &= \int_{t_0}^{t_f} \Big(\int_{t}^{t_f} \lambda(s)ds\Big)^T B\Big(t,\omega,v(\cdot)\Big)h(t)dt = \int_{t_0}^{t_f} \Big(B\Big(t,\omega,v(\cdot)\Big)^T \int_{t}^{t_f} \lambda(s)ds\Big)^T h(t)dt. \end{aligned} \tag{3.22b}$$

Inserting (3.22b) into (3.20c), we have

$$\begin{aligned} F'_{v(\cdot)+}\Big(u(\cdot);h(\cdot)\Big) &= E\Big(\int_{t_0}^{t_f} \Big(B\Big(t,\omega,v(\cdot)\Big)^T \int_{t}^{t_f} \lambda(s)ds \\ &\quad + \tilde{b}\Big(t,\omega,v(\cdot),u(\cdot)\Big)\Big)^T h(t)dt\Big). \end{aligned} \tag{3.23}$$

By means of (3.20d), the integral equation (3.21) may be written by

$$\lambda(t) - A\Big(t,\omega,v(\cdot)\Big)^T \Big(c\Big(t_f,\omega,v(\cdot),u(\cdot)\Big) + \int_{t}^{t_f} \lambda(s)ds\Big) = a\Big(t,\omega,v(\cdot),u(\cdot)\Big). \tag{3.24}$$

According to (3.24), defining the m-vector function

$$y = y_{v,u}(t,\omega) := c\Big(t_f,\omega,v(\cdot),u(\cdot)\Big) + \int_{t}^{t_f} \lambda_{v,u}(s,\omega)ds, \tag{3.25a}$$

we get

$$\lambda(t) - A\Big(t,\omega,v(\cdot)\Big)^T y_{v,u}(t,\omega) = a\Big(t,\omega,v(\cdot),u(\cdot)\Big). \tag{3.25b}$$

Replacing in (3.25b) the variable t by s and integrating then the equation (3.25b) over the time interval $[t,t_f]$, yields

$$\int_t^{t_f} \lambda(s)ds = \int_t^{t_f} \Big(A\Big(s, \omega, v(\cdot)\Big)^T y_{v,u}(s, \omega) + a\Big(s, \omega, v(\cdot), u(\cdot)\Big)\Big) ds. \quad (3.25c)$$

Finally, using again (3.25a), from (3.25c) we get

$$y_{v,u}(t, \omega) = c\Big(t_f, \omega, v(\cdot), u(\cdot)\Big) + \int_t^{t_f} \Big(A\Big(s, \omega, v(\cdot)\Big)^T y_{v,u}(s, \omega) + a\Big(s, \omega, v(\cdot), u(\cdot)\Big)\Big) ds. \quad (3.25d)$$

Obviously, the differential form of the Volterra integral equation (3.25d) for $y = y_{v,u}(t, \omega)$ reads

$$\dot{y}(t) = -A\Big(t, \omega, v(\cdot)\Big)^T y(t) - a\Big(t, \omega, v(\cdot), u(\cdot)\Big), t_0 \le t \le t_f, \quad (3.26a)$$

$$y(t_f) = c\Big(t_f, \omega, v(\cdot), u(\cdot)\Big). \quad (3.26b)$$

System (3.25d), (3.26a), (3.26b), resp., is called the *adjoint integral, differential equation* related to the perturbation equation (3.16a), (3.16b).

By means of (3.25a), from (3.20e) and (3.23) we now obtain

$$\begin{aligned} F'_{v(\cdot)+}\Big(u(\cdot); h(\cdot)\Big) &= E \int_{t_0}^{t_f} \Big(B\Big(t, \omega, v(\cdot)\Big)^T \int_t^{t_f} \lambda(s)ds + b\Big(t, \omega, v(\cdot), u(\cdot)\Big) \\ &\quad + B\Big(t, \omega, v(\cdot)\Big)^T c\Big(t_f, \omega, v(\cdot), u(\cdot)\Big)\Big)^T h(t)dt \\ &= E \int_{t_0}^{t_f} \Big(B\Big(t, \omega, v(\cdot)\Big)^T y_{v,u}(t, \omega) + b\Big(t, \omega, v(\cdot), u(\cdot)\Big)\Big)^T h(t)dt. \end{aligned}$$

Summarizing the above transformations, we have the following result.

Theorem 3.3 *Let $(\omega, t) \to y_{v,u}(t, \omega)$ be the unique measurable solution of the adjoint integral, differential equation (3.25d), (3.26a), (3.26b), respectively. Then,*

$$\begin{aligned} F'_{v(\cdot)+}\Big(u(\cdot); h(\cdot)\Big) &= E \int_{t_0}^{t_f} \Big(B\Big(t, \omega, v(\cdot)\Big)^T y_{v,u}(t, \omega) \\ &\quad + b\Big(t, \omega, v(\cdot), u(\cdot)\Big)\Big)^T h(t)dt. \end{aligned} \quad (3.27)$$

Note that $F'_{v(\cdot)+}\Big(u(\cdot);\cdot\Big)$ is also the Gâteaux-differential of $F_{v(\cdot)}$ at $u(\cdot)$.

For a further discussion of formula (3.27) for $F'_{v(\cdot)+}\Big(u(\cdot);h(\cdot)\Big)$, in generalization of the *Hamiltonian* of a deterministic control problem, see, e.g., [18], we introduce now the **stochastic Hamiltonian** related to the partly linearized control problem (3.17a)–(3.17d) based on a reference control $v(\cdot)$:

$$H_{v(\cdot)}(t,\omega,\zeta,y,u) := L\Big(t,\omega,z_v(t,\omega)+\zeta,u\Big)$$
$$+y^T\Big(A\Big(t,\omega,v(\cdot)\Big)\zeta + B\Big(t,\omega,v(\cdot)\Big)\Big(u-v(t)\Big)\Big), \tag{3.28a}$$

$(t,\omega,z,y,u)\in[t_0,t_f]\times\Omega\times\mathbb{R}^m\times\mathbb{R}^m\times\mathbb{R}^n$. Using $H_{v(\cdot)}$, for $F'_{v(\cdot)+}$ we find the representation

$$F'_{v(\cdot)+}\Big(u(\cdot);h(\cdot)\Big) = \int_{t_0}^{t_f} E\nabla_u H_{v(\cdot)}\Big(t,\omega,\zeta_{v,u-v}(t,\omega),$$
$$y_{v,u}(t,\omega),u(t)\Big)^T h(t)dt, \tag{3.28b}$$

where $\zeta_{v,u-v}=\zeta_{v,u-v}(t,\omega)$ is the solution of the perturbation differential, integral equation (3.17b), (3.17c), (3.20b), resp., and $y_{v,u}=y_{v,u}(t,\omega)$ denotes the solution of the adjoint differential, integral equation (3.26a), (3.26b), (3.25d).

Let $u^0(\cdot)\in U$ denote a given initial control. By means of (3.28a), (3.28b), the necessary and sufficient condition for a control $u^1(\cdot)$ to be an element of the set $M\Big(u^0(\cdot)\Big)$, i.e., a solution of the approximate convex problem $(3.7)_{u^0(\cdot)}$, reads, see Definition 3.4 and (3.13):

$$\int_{t_0}^{t_f} E\nabla_u H_{u^0(\cdot)}\Big(t,\omega,\zeta_{u^0,u^1-u^0}(t,\omega),y_{u^0,u^1}(t,\omega),$$
$$u^1(t)\Big)^T\Big(u(t)-u^1(t)\Big)dt\geq 0, u(\cdot)\in D. \tag{3.29}$$

Introducing, for given controls $u^0(\cdot),u^1(\cdot)$, the convex mean value function $\overline{H}_{u^0(\cdot),u^1(\cdot)}=\overline{H}_{u^0(\cdot),u^1(\cdot)}(u(\cdot))$ defined by

$$\overline{H}_{u^0(\cdot),u^1(\cdot)}\Big(u(\cdot)\Big) := \int_{t_0}^{t_f} EH_{u^0(\cdot)}\Big(t,\omega,\zeta_{u^0,u^1-u^0}(t,\omega),$$
$$y_{u^0,u^1}(t,\omega),u(t)\Big)dt, \tag{3.30a}$$

corresponding to the representation (3.12a) and (3.18) of the directional derivative of F, it is seen that the left hand side of (3.29) is the directional derivative of the function $\overline{H}_{u^0(\cdot),u^1(\cdot)} = \overline{H}_{u^0(\cdot),u^1(\cdot)}(u(\cdot))$ at $u^1(\cdot)$ with increment $h(\cdot) := u(\cdot) - u^1(\cdot)$. Consequently, (3.29) is equivalent with the condition:

$$\overline{H}'_{u^0(\cdot),u^1(\cdot)+}\Big(u^1(\cdot)); \big(u(\cdot) - u^1(\cdot)\big) \geq 0, u(\cdot) \in D. \tag{3.30b}$$

Due to the equivalence of the conditions (3.29) and (3.30b), for a control $u^1(\cdot) \in M\Big(u^0(\cdot)\Big)$, i.e., a solution of $(3.7)_{u^0(\cdot)}$ we have the following characterization:

Theorem 3.4 *Let $u^0(\cdot) \in D$ be a given initial control. A control $u^{(1)}(\cdot) \in U$ is a solution of $(3.7)_{u^0(\cdot)}$, i.e., $u^1(\cdot) \in M\Big(u^0(\cdot)\Big)$, if and only if $u^1(\cdot)$ is an optimal solution of the convex stochastic optimization problem*

$$\min \overline{H}_{u^0(\cdot),u^1(\cdot)}\Big(u(\cdot)\Big) \quad s.t.\ u(\cdot) \in D. \tag{3.31a}$$

In the following we study therefore the convex optimization problem (3.31a), where we replace next to the yet unknown functions $\zeta = \zeta_{u^0,u^1-u^0}(t,\omega)$, $y = y_{u^0,u^1}(t,\omega)$ by arbitrary stochastic functions $\zeta = \zeta(t,\omega)$, $y = y(t,\omega)$. Hence, we consider the mean value function $\overline{H}_{u^0(\cdot),\zeta(\cdot,\cdot),y(\cdot,\cdot)} = \overline{H}_{u^0(\cdot),\zeta(\cdot,\cdot),y(\cdot,\cdot)}(u(\cdot))$ defined, see (3.30a), by

$$\overline{H}_{u^0(\cdot),\zeta(\cdot,\cdot),y(\cdot,\cdot)}\Big(u(\cdot)\Big) = \int_{t_0}^{t_f} E\,H_{u^0(\cdot)}\Big(t,\omega,\zeta(t,\omega),y(t,\omega),u(t)\Big)dt. \tag{3.31b}$$

In practice, the admissible domain D is often given by

$$D = \Big\{u(\cdot) \in U : u(t) \in D_t, t_0 \leq t \leq t_f\Big\}, \tag{3.32}$$

where $D_t \subset \mathbb{R}^n$ is a given convex subset of $\mathbb{R}^n$ for each time t, $t_0 \leq t \leq t_f$. Since $\overline{H}_{u^0(\cdot),\zeta(\cdot,\cdot),y(\cdot,\cdot)}\Big(u(\cdot)\Big)$ has an integral form, the minimum value of $\overline{H}_{u^0(\cdot),\zeta(\cdot,\cdot),y(\cdot,\cdot)}\Big(u(\cdot)\Big)$ on D can be obtained—in case (3.32)—by solving the finite-dimensional stochastic optimization problem

$$\min E\,H_{u^0(\cdot)}\Big(t,\omega,\zeta(t,\omega), y(t,\omega),u\Big) \text{ s.t. } u \in D_t \tag{P}^t_{u^0(\cdot),\zeta,y}$$

for each $t \in [t_0, t_f]$. Let denote then

$$\widetilde{u}^* = \widetilde{u}^*_{u^0(\cdot)}\Big(t, \zeta(t,\cdot), y(t,\cdot)\Big), \quad t_0 \le t \le t_f, \tag{3.33a}$$

a solution of $(P)^t_{u^0(\cdot),\zeta,y}$ for each $t_0 \le t \le t_f$. Obviously, if

$$\widetilde{u}^*_{u^0(\cdot)}\Big(\cdot, \zeta(\cdot,\cdot), y(\cdot,\cdot)\Big) \in U\Big(\text{ and therefore } \widetilde{u}^*_{u^0(\cdot)}\Big(\cdot, \zeta(\cdot,\cdot), y(\cdot,\cdot)\Big) \in D\Big), \tag{3.33b}$$

then

$$\widetilde{u}^*_{u^0(\cdot)}\Big(\cdot, \zeta(\cdot,\cdot), y(\cdot,\cdot)\Big) \in \underset{u(\cdot)\in D}{\operatorname{argmin}}\, \overline{H}_{u^0(\cdot),\zeta(\cdot,\cdot),y(\cdot,\cdot)}\Big(u(\cdot)\Big). \tag{3.33c}$$

Because of Theorem 3.4, problems $(P)^t_{u^0(\cdot),\zeta,y}$ and (3.33a)–(3.33c), we introduce, cf. [18], the following definition:

Definition 3.6 Let $u^0(\cdot) \in D$ be a given initial control. For measurable functions $\zeta = \zeta(t,\omega)$, $y = y(t,\omega)$ on $(\Omega, \mathfrak{A}, P)$, let denote

$$\widetilde{u}^* = \widetilde{u}^*_{u^0(\cdot)}\Big(t, \zeta(t,\cdot), y(t,\cdot)\Big), \quad t_0 \le t \le t_f,$$

a solution of $(P)^t_{u^0(\cdot),\zeta,y}$, $t_0 \le t \le t_f$. The function $\widetilde{u}^* = \widetilde{u}^*_{u^0(\cdot)}\Big(t, \zeta(t,\cdot), y(t,\cdot)\Big)$ is called a $\mathbf{H_{u^0(\cdot)}}$**-minimal control** of (3.17a)–(3.17d). The stochastic Hamiltonian $H_{u^0(\cdot)}$ is called **regular, strictly regular**, resp., if a $H_{u^0(\cdot)}$-minimal control exists and is determined uniquely.

Remark 3.10 Obviously, by means of Definition 3.6, for an optimal solution $u^1(\cdot)$ of $(3.7)_{u^0(\cdot)}$ we have then the "*model control law*" $\widetilde{u}^* = \widetilde{u}^*_{u^0(\cdot)}\Big(t, \zeta(t,\cdot), y(t,\cdot)\Big)$ depending on still unknown state, costate functions $\zeta(t,\cdot)$, $y(\cdot)$, respectively.

3.5 Canonical (Hamiltonian) System of Differential Equations/Two-Point Boundary Value Problem

For a given initial control $u^0(\cdot) \in D$, let $\widetilde{u}^* = \widetilde{u}^*_{u^0(\cdot)}\Big(t, \zeta(\cdot), \eta(\cdot)\Big)$, $t_0 \le t \le t_f$, denote a $H_{u^0(\cdot)}$-minimal control of (3.17a)–(3.17d) according to Definition 3.6. Due to (3.17b), (3.17c) and (3.26a), (3.26b) we consider, cf. [18], the following so-called canonical or Hamiltonian system of differential equations, hence, a two-point boundary value problem, with random parameters for the vector functions $(\zeta, y) = \Big(\zeta(t,\omega), y(t,\omega)\Big)$, $t_0 \le t \le t_f$, $\omega \in \Omega$:

$$\dot{\zeta}(t,\omega) = A\Big(t,\omega,u^0(\cdot)\Big)\zeta(t,\omega) + B\Big(t,\omega,u^0(\cdot)\Big)\Big(\widetilde{u}^*_{u^0(\cdot)}\Big(t,\zeta(t,\cdot),y(t,\cdot)\Big) - u^0(t)\Big),$$
$$t_0 \le t \le t_f, \tag{3.34a}$$
$$\zeta(t_0,\omega) = 0 \text{ a.s.} \tag{3.34b}$$
$$\dot{y}(t,\omega) = -A\Big(t,\omega,u^0(\cdot)\Big)^T y(t,\omega) - \nabla_z L\Big(t,\omega,z_{u^0}(t,\omega)+\zeta(t,\omega),\widetilde{u}^*_{u^0(\cdot)}\Big(t,\zeta(t,\cdot),y(t,\cdot)\Big)\Big),$$
$$t_0 \le t \le t_f, \tag{3.34c}$$
$$y(t_f,\omega) = \nabla_z G\Big(t_f,\omega,z_{u^0}(t_f,\omega)+\zeta(t_f,\omega)\Big). \tag{3.34d}$$

Remark 3.11 Note that the (deterministic) control law $\widetilde{u}^*=\widetilde{u}^*_{u^0(\cdot)}\Big(t,\zeta(t,\cdot),y(t,\cdot)\Big)$ depends on the whole random variable $\Big(\zeta(t,\omega),y(t,\omega)\Big)$, $\omega\in\Omega$, or the occurring moments. In the case of a discrete parameter distribution $\Omega=\{\omega_1,\ldots,\omega_\varrho\}$, see Sect. 3.1, the control law $\widetilde{u}^*_{u^0(\cdot)}$ depends,

$$\widetilde{u}^*_{u^0(\cdot)} = \widetilde{u}^*_{u^0(\cdot)}\Big(t,\zeta(t,\omega_1),\ldots,\zeta(t,\omega_\varrho),y(t,\omega_1),\ldots,y(t,\omega_\varrho)\Big),$$

on the 2ϱ unknown functions $\zeta(t,\omega_i)$, $y(t,\omega_i)$, $i=1,\ldots,\varrho$.

Suppose now that $(\zeta^1,y^1)=\Big(\zeta^1(t,\omega),y^1(t,\omega)\Big)$, $t_0\le t\le t_f$, $\omega\in(\Omega,\mathfrak{A},P)$, is the unique measurable solution of the canonical stochastic system (3.34a)–(3.34d), and define

$$u^1(t) := \widetilde{u}^*_{u^0(\cdot)}\Big(t,\zeta^1(t,\cdot),y^1(t,\cdot)\Big),\ t_0\le t\le t_f. \tag{3.35}$$

System (3.34a)–(3.34d) takes then the following form:

$$\dot{\zeta}^1(t,\omega) = A\Big(t,\omega,u^0(\cdot)\Big)\zeta^1(t,\omega) + B\Big(t,\omega,u^0(\cdot)\Big)\Big(u^1(t)-u^0(t)\Big),$$
$$t_0\le t\le t_f, \tag{3.34a$'$}$$
$$\zeta^1(t_0,\omega) = 0 \text{ a.s.} \tag{3.34b$'$}$$
$$\dot{y}^1(t,\omega) = -A\Big(t,\omega,u^0(\cdot)\Big)^T y(t,\omega) - \nabla_z L\Big(t,\omega,z_{u^0}(t,\omega)+\zeta^1(t,\omega),u^1(t)\Big),\ t_0\le t\le t_f, \tag{3.34c$'$}$$
$$y^1(t_f,\omega) = \nabla_z G\Big(t_f,\omega,z_{u^0}(t_f,\omega)+\zeta^1(t_f,\omega)\Big). \tag{3.34d$'$}$$

Assuming that

$$u^1(\cdot)\in U, \tag{3.36a}$$

due to the definition of a $H_{u^0(\cdot)}$-minimal control we also have

$$u^1(\cdot) \in D. \tag{3.36b}$$

According to the notation introduced in (3.16a)–(3.16g), (3.25a)–(3.25d)/(3.26a), (3.26b), resp., and the above-assumed uniqueness of the solution (ζ^1, y^1) of (3.34a)–(3.34d) we have

$$\zeta^1(t,\omega) = \zeta_{u^0,u^1-u^0}(t,\omega) \tag{3.37a}$$
$$y^1(t,\omega) = y_{u^0,u^1}(t,\omega) \tag{3.37b}$$

with the control $u^1(\cdot)$ given by (3.35).

Due to the above construction, we know that $u^1(t)$ solves $(P)^t_{u^0(\cdot),\zeta,\eta}$ for

$$\zeta = \zeta(t,\omega) := \zeta^1(t,\omega) = \zeta_{u^0,u^1-u^0}(t,\omega)$$
$$\eta = \eta(t,\omega) := y^1(t,\omega) = y_{u^0,u^1}(t,\omega)$$

for each $t_0 \le t \le t_f$. Hence, control $u^1(\cdot)$, given by (3.35), is a solution of (3.31a).

Summarizing the above construction, from Theorem 3.4 we obtain this result.

Theorem 3.5 *Suppose that D is given by (3.32), $M\left(u^0(\cdot)\right) \neq \emptyset$, and $(P)^t_{u^0(\cdot),\zeta,y}$ has an optimal solution for each t, $t_0 \le t \le t_f$, and measurable functions $\zeta(t,\cdot)$, $y(t,\cdot)$. Moreover, suppose that the canonical system (3.34a)–(3.34d) has a unique measurable solution $\left(\zeta^1(t,\omega), y^1(t,\omega)\right)$, $t_0 \le t \le t_f$, $\omega \in \Omega$, such that $u^1(\cdot) \in U$, where $u^1(\cdot)$ is defined by (3.35). Then $u^1(\cdot)$ is a solution of* $(3.7)_{u^0(\cdot)}$.

3.6 Stationary Controls

Suppose here that condition (3.14) holds for all controls $v(\cdot)$ under consideration, cf. Lemma 3.4. Having a method for the construction of improved approximative controls $u^1(\cdot) \in M\left(u^0(\cdot)\right)$ related to an initial control $u^0(\cdot) \in D$, we consider now the construction of stationary controls of the control problem (3.7), i.e., elements $\bar{u}(\cdot) \in D$ such that $\bar{u}(\cdot) \in M\left(\bar{u}(\cdot)\right)$, see Definition 3.5.

Starting again with formula (3.27), by means of (3.14), for an element $v(\cdot) \in D$ we have

$$F'_+\Big(v(\cdot);h(\cdot)\Big) = F'_{v(\cdot)+}\Big(v(\cdot);h(\cdot)\Big)$$
$$= \int_{t_0}^{t_f} E\bigg(B\Big(t,\omega,v(\cdot)\Big)^T y_v(t,\omega) + b\Big(t,\omega,v(\cdot),v(\cdot)\Big)\bigg)^T h(t)dt, \quad (3.38a)$$

where

$$y_v(t,\omega) := y_{v,v}(t,\omega) \quad (3.38b)$$

fulfills, cf. (3.26a), (3.26b), the adjoint differential equation

$$\dot{y}(t,\omega) = -A\Big(t,\omega,v(\cdot)\Big)^T y(t,\omega) - \nabla_z L\Big(t,\omega,z_v(t,\omega),v(t)\Big) \quad (3.39a)$$
$$y(t_f,\omega) = \nabla_z G\Big(t_f,\omega,z_v(t_f,\omega)\Big). \quad (3.39b)$$

Moreover, cf. (3.16b), (3.16c),

$$A\Big(t,\omega,v(\cdot)\Big) = D_z g\Big(t,\omega,z_v(t,\omega),v(t)\Big) \quad (3.39c)$$
$$B\Big(t,\omega,v(\cdot)\Big) = D_u g\Big(t,\omega,z_v(t,\omega),v(t)\Big) \quad (3.39d)$$

and, see (3.19b),

$$b\Big(t,\omega,v(\cdot),v(\cdot)\Big) = \nabla_u L\Big(t,\omega,z_v(t,\omega),v(t)\Big), \quad (3.39e)$$

where $z_v = z_v(t,\omega)$ solves the dynamic equation

$$\dot{z}(t,\omega) = g\Big(t,\omega,z(t,\omega),v(t)\Big), t_0 \le t \le t_f, \quad (3.39f)$$
$$z(t_0,\omega) = z_0(\omega). \quad (3.39g)$$

Using now the stochastic Hamiltonian, cf. (3.28a),

$$H(t,\omega,z,y,u) := L(t,\omega,z,u) + y^T g(t,\omega,z,u) \quad (3.40a)$$

related to the basic control problem (3.9a)–(3.9d), from (3.38a), (3.38b), (3.39a)–(3.39g) we get the representation, cf. (3.28a), (3.28b),

$$F'_+\Big(v(\cdot);h(\cdot)\Big) = \int_{t_0}^{t_f} E\nabla_u H\Big(t,\omega,z_v(t,\omega),y_v(t,\omega),v(t)\Big)^T h(t)dt. \quad (3.40b)$$

According to condition (3.13), a stationary control of (3.7), hence, an element $\bar{u}(\cdot) \in D$ such that $\bar{u}(\cdot)$ is an optimal solution of $(3.7)_{\bar{u}(\cdot)}$ is characterized, see (3.14), by

$$F'_+\Big(\bar{u}(\cdot); u(\cdot) - \bar{u}(\cdot)\Big) \geq 0 \text{ for all } u(\cdot) \in D.$$

Thus, for stationary controls $\bar{u}(\cdot) \in D$ of problem (3.7) we have the characterization

$$\int_{t_0}^{t_f} E\nabla_u H\Big(t, \omega, z_{\bar{u}}(t, \omega), y_{\bar{u}}(t, \omega), \bar{u}(t)\Big)^T \Big(u(t) - \bar{u}(t)\Big)dt \geq 0, u(\cdot) \in D. \quad (3.41)$$

Comparing (3.29) and (3.41), corresponding to (3.30a), (3.30b), for given $w(\cdot) \in D$ we introduce here the function

$$\overline{H}_w\Big(u(\cdot)\Big) := \int_{t_0}^{t_f} EH\Big(t, \omega, z_w(t, \omega), y_w(t, \omega), u(t)\Big)dt, \quad (3.42a)$$

and we consider the optimization problem

$$\min \overline{H}_w\Big(u(\cdot)\Big) \text{ s.t. } u(\cdot) \in D. \quad (3.42b)$$

Remark 3.12 Because of the assumptions in Sect. 3.1.2, problem (3.42b) is (strictly) convex, provided that the process differential equation (3.1a) is affine-linear with respect to u, hence,

$$\dot{z}(t, \omega) = g(t, \omega, z, u) = \hat{g}(t, \omega, z) + \hat{B}(t, \omega, z)u \quad (3.43)$$

with a given vector-, matrix-valued function $\hat{g} = \hat{g}(t, \omega, z)$, $\hat{B} = \hat{B}(t, \omega, z)$.

If differentiation and integration/expectation in (3.42a) may be interchanged, which is assumed in the following, then (3.41) is a necessary condition for

$$\bar{u}(\cdot) \in \operatorname*{argmin}_{u(\cdot) \in D} \overline{H}_{\bar{u}}\Big(u(\cdot)\Big), \quad (3.44)$$

cf. (3.30a), (3.30b), (3.31a), (3.31b). Corresponding to Theorem 3.4, here we have this result:

Theorem 3.6 *(*Optimality condition for stationary controls*) Suppose that a control $\bar{u}(\cdot) \in D$ fulfills (3.44). Then $\bar{u}(\cdot)$ is a stationary control of (3.7).*

3.7 Canonical (Hamiltonian) System of Differential

Assume now again that the feasible domain D is given by (3.32). In order to solve the optimization problem (3.42a), (3.42b), corresponding to $(P)^t_{u^0(\cdot),\zeta,\eta}$, here we consider the finite-dimensional optimization problem

$$\min E H\Big(t, \omega, \zeta(\omega), \eta(\omega), u\Big) \quad \text{s.t. } u \in D_t \qquad (P)^t_{\zeta,\eta}$$

for each t, $t_0 \le t \le t_f$. Furthermore, we use again the following definition, cf. Definition 3.6.

Definition 3.7 For measurable functions $\zeta(\cdot)$, $\eta(\cdot)$ on $(\Omega, \mathcal{A}, P)$, let denote

$$\widetilde{u^*} = \widetilde{u^*}\Big(t, \zeta(\cdot), \eta(\cdot)\Big), t_0 \le t \le t_f,$$

a solution of $(P)^t_{\zeta,\eta}$. The function $\widetilde{u^*} = \widetilde{u^*}\Big(t, \zeta(\cdot), \eta(\cdot)\Big), t_0 \le t \le t_f$, is called a **H-minimal control** of (3.9a)–(3.9d). The stochastic Hamiltonian H is called **regular, strictly regular**, resp., if a H-minimal control exists, exists and is determined uniquely.

For a given H-minimal control $u^* = u^*\Big(t, \zeta(\cdot), \eta(\cdot)\Big)$ we consider now, see (3.34a)–(3.34d), the following canonical (Hamiltonian) two-point boundary value problem with random parameters:

$$\dot{z}(t, \omega) = g\Big(t, \omega, z(t, \omega), u^*\Big(t, z(t, \cdot), y(t, \cdot)\Big)\Big), t_0 \le t \le t_f \qquad (3.45a)$$

$$z(t_0, \omega) = z_0(\omega) \qquad (3.45b)$$

$$\dot{y}(t, \omega) = -D_z g\Big(t, \omega, z(t, \omega), u^*\Big(t, z(t, \cdot), y(t, \cdot)\Big)\Big)^T y(t, \omega)$$

$$-\nabla_z L\Big(t, \omega, z(t, \omega), u^*\Big(t, z(t, \cdot), y(t, \cdot)\Big)\Big), t_0 \le t \le t_f \qquad (3.45c)$$

$$y(t_f, \omega) = \nabla_z G\Big(t_f, \omega, z(t_f, \omega)\Big). \qquad (3.45d)$$

Remark 3.13 In case of a discrete distribution $\Omega = \{\omega_1, \ldots, \omega_\varrho\}$, $P(\omega = \omega_j)$, $j = 1, \ldots, \varrho$, corresponding to Sect. 3.1, for the H-minimal control we have

$$\widetilde{u^*} = \widetilde{u^*}\Big(t, z(t, \omega_1), \ldots, z(t, \omega_\varrho), y(t, \omega_1), \ldots, y(t, \omega_\varrho)\Big).$$

Thus, (3.45a)–(3.45d) is then an ordinary two-point boundary value problem for the 2ϱ unknown functions $z = z(t, \omega_j)$, $y = y(t, \omega_j)$, $j = 1, \ldots, \varrho$.

Let denote $(\bar{z}, \bar{y}) = \Big(\bar{z}(t,\omega), \bar{y}(t,\omega)\Big)$, $t_0 \le t \le t_f$, $\omega \in (\Omega, \mathfrak{A}, P)$, the unique measurable solution of (3.45a)–(3.45d) and define:

$$\bar{u}(t) := \widetilde{u}^*\Big(t, \bar{z}(t,\cdot), \bar{y}(t,\cdot)\Big), \quad t_0 \le t \le t_f. \tag{3.46}$$

Due to (3.16e) and (3.38b), (3.39a), (3.39b) we have

$$\bar{z}(t,\omega) = z_{\bar{u}}(t,\omega), \quad t_0 \le t \le t_f, \ \omega \in (\Omega, \mathfrak{A}, P) \tag{3.47a}$$

$$\bar{y}(t,\omega) = y_{\bar{u}}(t,\omega), \quad t_0 \le t \le t_f, \ \omega \in (\Omega, \mathcal{A}, P), \tag{3.47b}$$

hence,

$$\bar{u}(t) = \widetilde{u}^*\Big(t, z_{\bar{u}}(t,\cdot), y_{\bar{u}}(t,\cdot)\Big), \quad t_0 \le t \le t_f. \tag{3.47c}$$

Assuming that

$$\bar{u}(\cdot) \in U \quad \text{(and therefore } \bar{u}(\cdot) \in D\text{)}, \tag{3.48}$$

we get this result:

Theorem 3.7 *Suppose that the Hamiltonian system* (3.45a)–(3.45d) *has a unique measurable solution* $(\bar{z}, \bar{y}) = \Big(\bar{z}(t,\omega), \bar{y}(t,\omega)\Big)$*, and define* $\bar{u}(\cdot)$ *by* (3.46) *with a* H*-minimal control* $\widetilde{u}^* = \widetilde{u}^*(t,\zeta,\eta)$*. If* $\bar{u}(\cdot) \in U$*, then* $\bar{u}(\cdot)$ *is a stationary control.*

Proof According to the construction of $(\bar{z}, \bar{y}, \bar{u})$, the control $\bar{u}(\cdot) \in D$ minimizes $\overline{H}_{\bar{u}}\Big(u(\cdot)\Big)$ on D. Hence,

$$\bar{u}(\cdot) \in \operatorname*{argmin}_{u(\cdot) \in D} \overline{H}_{\bar{u}}\Big(u(\cdot)\Big).$$

Theorem 3.6 yields then that $\bar{u}(\cdot)$ is a stationary control. □

3.8 Computation of Expectations by Means of Taylor Expansions

Corresponding to the assumptions in Sect. 3.1, based on a parametric representation of the random differential equation with a finite-dimensional random parameter vector $\theta = \theta(\omega)$, we suppose that

$$g(t, w, z, u) = \tilde{g}(t, \theta, z, u) \tag{3.49a}$$
$$z_0(\omega) = \tilde{z}_0(\theta) \tag{3.49b}$$
$$L(t, \omega, z, u) = \tilde{L}(t, \theta, z, u) \tag{3.49c}$$
$$G(t, \omega, z) = \tilde{G}(t, \theta, z). \tag{3.49d}$$

Here,

$$\theta = \theta(\omega), \omega \in (\Omega, \mathfrak{A}, P) \tag{3.49e}$$

denotes the time-independent r-vector of random model parameters and random initial values, and $\tilde{g}$, $\tilde{z}_0$, $\tilde{L}$, $\tilde{G}$ are sufficiently smooth functions of the variables indicated in (3.49a)–(3.49d). For simplification of notation we omit symbol "$\sim$" and write

$$g(t, w, z, u) := g\Big(t, \theta(\omega), z, u\Big) \tag{3.49a$'$}$$
$$z_0(\omega) := z_0\Big(\theta(\omega)\Big) \tag{3.49b$'$}$$
$$L(t, \omega, z, u) := L\Big(t, \theta(\omega), z, u\Big) \tag{3.49c$'$}$$
$$G(t, \omega, z, u) := G\Big(t, \theta(\omega), z\Big). \tag{3.49d$'$}$$

Since the approximate problem (3.17a)–(3.17d), obtained by the above described *inner linearization*, has the same basic structure as the original problem (3.9a)–(3.9d), it is sufficient to describe the procedure for problem (3.9a)–(3.9d). Again, for simplification, the conditional expectation $E(\ldots|\mathfrak{A}_{t_0})$ given the information $\mathfrak{A}_{t_0}$ up to the considered starting time t_0 is denoted by "E". Thus, let denote

$$\overline{\theta} = \overline{\theta}^{t_0} := E\theta(\omega) = E\Big(\theta(\omega)|\mathfrak{A}_{t_0}\Big) \tag{3.50a}$$

the conditional expectation of the random vector $\theta(\omega)$ given the information $\mathfrak{A}_{t_0}$ at time point t_0. Taking into account the properties of the solution

$$z = z(t, \theta) = S\Big(z_0(\theta), \theta, u(\cdot)\Big)(t), t \geq t_0, \tag{3.50b}$$

of the dynamic equation (3.3a)–(3.3d), see Lemma 3.1, the expectations arising in the objective function (3.9a) can be computed approximatively by means of Taylor expansion with respect to θ at $\overline{\theta}$.

3.8.1 Complete Taylor Expansion

Considering first the costs L along the trajectory we obtain, cf. [26],

$$\begin{aligned} L\Big(t,\theta,z(t,\theta),u(t)\Big) &= L\Big(t,\overline{\theta},z(t,\overline{\theta}),u(t)\Big) \\ &+\Big(\nabla_\theta L\Big(t,\overline{\theta},z(t,\overline{\theta}),u(t)\Big)+D_\theta z(t,\overline{\theta})^T\nabla_z L\Big(t,\overline{\theta},z(t,\overline{\theta}),u(t)\Big)\Big)^T(\theta-\overline{\theta}) \\ &+\frac{1}{2}(\theta-\overline{\theta})^T Q_L\Big(t,\overline{\theta},z(t,\overline{\theta}),D_\theta z(t,\overline{\theta}),u(t)\Big)(\theta-\overline{\theta})+\dots\,. \end{aligned} \tag{3.51a}$$

Retaining only first-order derivatives of $z=z(t,\theta)$ with respect to θ, the approximative Hessian Q_L of $\theta\to L\Big(t,\theta,z(t,\theta),u\Big)$ at $\theta=\overline{\theta}$ is given by

$$\begin{aligned} Q_L\Big(t,\overline{\theta},z(t,\overline{\theta}),D_\theta z(t,\overline{\theta}),u(t)\Big) &:= \nabla_\theta^2 L\Big(t,\overline{\theta},z(t,\overline{\theta}),u(t)\Big) \\ &+D_\theta z(t,\overline{\theta})^T\nabla_{\theta z}^2 L\Big(t,\overline{\theta},z(t,\overline{\theta}),u(t)\Big)+\nabla_{\theta z}^2 L\Big(t,\overline{\theta},z(t,\overline{\theta}),u(t)\Big)^T D_\theta z(t,\overline{\theta}) \\ &+D_\theta z(t,\overline{\theta})^T\nabla_z^2 L\Big(t,\overline{\theta},z(t,\overline{\theta}),u(t)\Big)D_\theta z(t,\overline{\theta}). \end{aligned} \tag{3.51b}$$

Here, $\nabla_\theta L$, $\nabla_z L$ denotes the gradient of L with respect to θ, z, resp., $D_\theta z$ is the Jacobian of $z=z(t,\theta)$ with respect to θ, and $\nabla_\theta^2 L$, $\nabla_z^2 L$, resp., denotes the Hessian of L with respect to θ, z. Moreover, $\nabla_{\theta z}^2 L$ is the $r\times m$ matrix of partial derivatives of L with respect to θ_i and z_k, in this order.

Taking expectations in (3.51a), from (3.51b) we obtain the expansion

$$\begin{aligned} EL\Big(t,\theta(\omega),z\Big(t,\theta(\omega)\Big),u(t)\Big) &= L\Big(t,\overline{\theta},z(t,\overline{\theta}),u(t)\Big) \\ &+\frac{1}{2}E\Big(\theta(\omega)-\overline{\theta}\Big)^T Q_L\Big(t,\overline{\theta},z(t,\overline{\theta}),D_\theta z(t,\overline{\theta}),u(t)\Big)\Big(\theta(\omega)-\overline{\theta}\Big)+\dots \\ &= L\Big(t,\overline{\theta},z(t,\overline{\theta}),u(t)\Big)+\frac{1}{2}trQ_L\Big(t,\overline{\theta},z(t,\overline{\theta}),D_\theta z(t,\overline{\theta}),u(t)\Big)\mathrm{cov}\Big(\theta(\cdot)\Big)+\dots\,. \end{aligned} \tag{3.52}$$

For the terminal costs G, corresponding to the above expansion we find

$$\begin{aligned} G\Big(t_f,\theta,z(t_f,\theta)\Big) &= G\Big(t_f,\overline{\theta},z(t_f,\overline{\theta})\Big) \\ &+\Big(\nabla_\theta G\Big(t_f,\overline{\theta},z(t_f,\overline{\theta})\Big)+D_\theta z(t_f,\overline{\theta})^T\nabla_z G\Big(t_f,\overline{\theta},z(t_f,\overline{\theta})\Big)\Big)^T(\theta-\overline{\theta}) \\ &+\frac{1}{2}(\theta-\overline{\theta})^T Q_G\Big(t_f,\overline{\theta},z(t_f,\overline{\theta}),D_\theta z(t_f,\overline{\theta})\Big)(\theta-\overline{\theta})+\dots, \end{aligned} \tag{3.53a}$$

where Q_G is defined in the same way as Q_L, see (3.51a). Taking expectations with respect to $\theta(\omega)$, we get

$$EG\Big(t_f, \theta(\omega), z\Big(t_f, \theta(\omega)\Big)\Big) = G\Big(t_f, \overline{\theta}, z(t_f, \overline{\theta})\Big)$$
$$+\frac{1}{2} tr\, Q_G\Big(t_f, \overline{\theta}, z(t_f, \overline{\theta}), D_\theta z(t_f, \overline{\theta})\Big)\mathrm{cov}\Big(\theta(\cdot)\Big) + \dots . \tag{3.53b}$$

Note 3.4 Corresponding to Definition 3.1 and (3.50a), for the mean and covariance matrix of the random parameter vector $\theta = \theta(\omega)$ we have

$$\overline{\theta} = \overline{\theta}^{(t_0)} := E\Big(\theta(\omega)|\mathfrak{A}_{t_0}\Big)$$
$$\mathrm{cov}\Big(\theta(\cdot)\Big) = \mathrm{cov}^{(t_0)}\Big(\theta(\cdot)\Big) := E\Big(\Big(\theta(\omega) - \overline{\theta}^{(t_0)}\Big)\Big(\theta(\omega) - \overline{\theta}^{(t_0)}\Big)^T \Big|\mathfrak{A}_{t_0}\Big).$$

3.8.2 Inner or Partial Taylor Expansion

Instead of a complete expansion of L, G with respect to θ, appropriate approximations of the expected costs EL,EG, resp., may be obtained by the inner first-order approximation of the trajectory, hence,

$$L\Big(t, \theta, z(t, \theta), u(t)\Big) \approx L\Big(t, \theta, z(t, \overline{\theta}) + D_\theta z(t, \overline{\theta})(\theta - \overline{\theta}), u(t)\Big). \tag{3.54a}$$

Taking expectations in (3.54a), for the expected cost function we get the approximation

$$EL\Big(t, \theta, z(t, \theta), u(t)\Big)$$
$$\approx EL\Big(t, \theta(\omega), z(t, \overline{\theta}) + D_\theta z(t, \overline{\theta})(\theta(\omega) - \overline{\theta}), u(t)\Big). \tag{3.54b}$$

In many important cases, as, e.g., for cost functions L being quadratic with respect to the state variable z, the above expectation can be computed analytically. Moreover, if the cost function L is convex with respect to z, then the expected cost function EL is convex with respect to both, the state vector $z(t, \overline{\theta})$ and the Jacobian matrix of sensitivities $D_\theta z(t, \overline{\theta})$ evaluated at the mean parameter vector $\overline{\theta}$.

Having the approximate representations (3.52), (3.53b), (3.54b), resp., of the expectations occurring in the objective function (3.9a), we still have to compute the trajectory $t \to z(t, \overline{\theta})$, $t \geq t_0$, related to the mean parameter vector $\theta = \overline{\theta}$ and the sensitivities $t \to \frac{\partial z}{\partial \theta_i}(t, \overline{\theta})$, $i = 1, \dots, r$, $t \geq t_0$, of the state $z = z(t, \theta)$ with respect

to the parameters $\theta_i, i = 1, \ldots, r$, at $\theta = \overline{\theta}$. According to (3.3a), (3.3b) or (3.9b), (3.9c), for $z = z(t, \overline{\theta})$ we have the system of differential equations

$$\dot{z}(t, \overline{\theta}) = g\Big(t, \overline{\theta}, z(t, \overline{\theta}), u(t)\Big), \ t \geq t_0, \tag{3.55a}$$

$$z(t_0, \overline{\theta}) = z_0(\overline{\theta}). \tag{3.55b}$$

Moreover, assuming that differentiation with respect to $\theta_i, i = 1, \ldots, r$, and integration with respect to time t can be interchanged, see Lemma 3.1, from (3.3c) we obtain the following system of linear perturbation the differential equation for the Jacobian $D_\theta z(t, \overline{\theta}) = \Big(\frac{\partial z}{\partial \theta_1}(t, \overline{\theta}), \frac{\partial z}{\partial \theta_2}(t, \overline{\theta}), \ldots, \frac{\partial z}{\partial \theta_r}(t, \overline{\theta})\Big), t \geq t_0$:

$$\begin{aligned} \frac{d}{dt}\Big(D_\theta z(t, \overline{\theta})\Big) &= D_z g\Big(t, \overline{\theta}, z(t, \overline{\theta}), u(t)\Big) D_\theta z(t, \overline{\theta}) \\ &\quad + D_\theta g\Big(t, \overline{\theta}, z(t, \overline{\theta}), u(t)\Big), t \geq t_0, \end{aligned} \tag{3.56a}$$

$$D_\theta z(t_0, \overline{\theta}) = D_\theta z_0(\overline{\theta}). \tag{3.56b}$$

Note 3.5 Equations (3.56a), (3.56b) is closely related to the perturbation equation (3.16a), (3.16b) for representing the derivative $D_u z$ of z with respect to the control u. Moreover, the matrix differential equation (3.56a) can be decomposed into the following r differential equations for the columns $\frac{\partial z}{\partial \theta_j}(t, \overline{\theta}), j = 1, \ldots, r$:

$$\begin{aligned} \frac{d}{dt}\Big(\frac{\partial z}{\partial \theta_j}(t, \overline{\theta})\Big) &= \frac{\partial g}{\partial \theta_j}\Big(t, \overline{\theta}, z(t, \overline{\theta}), u(t)\Big) \frac{\partial z}{\partial \theta_j}(t, \overline{\theta}) \\ &\quad + \frac{\partial g}{\partial \theta_j}\Big(t, \overline{\theta}, z(t, \overline{\theta}), u(t)\Big), t \geq t_0, j = 1, \ldots, r. \end{aligned} \tag{3.56c}$$

Denoting by

$$\tilde{L} = \tilde{L}\Big(t, \theta, z(t, \overline{\theta}), D_\theta z(t, \overline{\theta}), u(t)\Big), \tag{3.57a}$$

$$\tilde{G} = \tilde{G}\Big(t_f, \theta, z(t_f, \overline{\theta}), D_\theta z(t_f, \overline{\theta})\Big), \tag{3.57b}$$

the approximation of the cost functions L, G by complete, partial Taylor expansion, for the optimal control problem under stochastic uncertainty (3.9a)–(3.9d) we now obtain the following approximation:

Theorem 3.8 *Suppose that differentiation with respect to the parameters $\theta_i, i = 1, \ldots, r$, and integration with respect to time t can be interchanged in (3.3c). Retaining only first-order derivatives of $z = z(t, \theta)$ with respect to θ, the optimal control*

problem under stochastic uncertainty (3.9a)–(3.9d) *can be approximated by the ordinary deterministic control problem:*

$$\min \int_{t_0}^{t_f} E\tilde{L}\Big(t, \theta(\omega), z(t,\overline{\theta}), D_\theta z(t,\overline{\theta}), u(t)\Big)dt$$
$$+E\tilde{G}\Big(t_f, \theta(\omega), z(t_f,\overline{\theta}, D_\theta z(t,\overline{\theta})\Big) \tag{3.58a}$$

subject to

$$\dot{z}(t,\overline{\theta}) = g\Big(t,\overline{\theta}, z(t,\overline{\theta}), u(t)\Big), t \geq t_0, \tag{3.58b}$$
$$z(t_0,\overline{\theta}) = z_0(\overline{\theta}) \tag{3.58c}$$
$$\frac{d}{dt}\Big(D_\theta z(t,\overline{\theta})\Big) = D_z g\Big(t,\overline{\theta}, z(t,\overline{\theta}), u(t)\Big) D_\theta z(t,\overline{\theta})$$
$$+D_\theta g\Big(t,\overline{\theta}, z(t,\overline{\theta}), u(t)\Big), t \geq t_0, \tag{3.58d}$$
$$D_\theta z(t_0,\overline{\theta}) = D_\theta z_0(\overline{\theta}) \tag{3.58e}$$
$$u(\cdot) \in D. \tag{3.58f}$$

Remark 3.14 Obviously, the trajectory of the above deterministic substitute control problem (3.58a)–(3.58f) of the original optimal control problem under stochastic uncertainty (3.9a)–(3.9d) can be represented by the $m(r+1)-$ vector function:

$$t \to \xi(t) := \begin{pmatrix} z(t,\overline{\theta}) \\ \frac{\partial z}{\partial \theta_1}(t,\overline{\theta}) \\ \vdots \\ \frac{\partial z}{\partial \theta_r}(t,\overline{\theta}) \end{pmatrix}, \quad t_0 \leq t \leq t_f. \tag{3.59}$$

Remark 3.15 Constraints of the expectation type (3.9f), i.e.,

$$Eh_{II}\Big(t, \theta(\omega), z\Big(t, \theta(\omega)\Big)\Big) \leq (=)0$$

can be evaluated as in (3.52) and (3.53b). This yields then deterministic constraints for the unknown functions $t \to z(t,\overline{\theta})$ and $t \to D_\theta z(t,\overline{\theta}), t \geq t_0$.

Remark 3.16 The expectations $EH_{u^\circ(\cdot)}, EH$ arising in $(P)^t_{u^\circ(\cdot),\zeta,\eta}$, $(P)^t_{\zeta,\eta}$, resp., can be determined approximatively as described above.

References

1. Allgöwer, F.E.A. (ed.): Nonlinear Model Predictive Control. Birkhäuser Verlag, Basel (2000)
2. Aoki, M.: Optimization of Stochastic Systems - Topics in Discrete-Time Systems. Academic, New York (1967)
3. Åström, K.: Introduction to Stochastic Control Theory. Elsevier (1970)
4. Barucha-Reid, A.: Random Integral Equations. Academic, New York (1972)
5. Bauer, H.: Wahrscheinlichkeitstheorie und Grundzüge der Masstheorie. Walter de Gruyter & Co., Berlin (1968)
6. Bunke, H.: Gewöhnliche Differentialgleichungen mit zufälligen Parametern. Akademie-Verlag, Berlin (1972)
7. Carathéodory, C.: Vorlesungen über reelle Funktionen. Teubner, Leipzig (1918)
8. Dieudonné, J.: Foundations of Modern Analysis. Academic, New York (1969)
9. Dreyfus, S.E., Lew, A.: The Art of Dynamic Programming. Academic, New York (1977)
10. Dreyfus, S.: Some types of optimal control of stochastic systems. J. SIAM Control **2**(1), 120–134 (1964)
11. Dreyfus, S.: Dynamic Programming and the Calculus of Variations. Academic, New York (1965)
12. Dullerud, G., Paganini, F.: A Course in Robust Control Theory. Springer, New York (2000)
13. Findeisen, R., et al.: Sampled–data nonlinear model predictive control for constrained continuous time systems. In: Tarbouriech, S., et al. (ed.) Adaptive Strategies in Control Systems With Input and Output Constraints, pp. 207–235. Springer, Berlin (2007)
14. Garcia, C.E., et al.: Model predictive control: theory and practice - a survey. Automatica **25**(3), 335–348 (1989). https://doi.org/10.1016/0005-1098(89)90002-2
15. Gessing, R., Jacobs, O.L.R.: On the equivalence between optimal stochastic control and open-loop feedback control. Int. J. Control **40**(1), 193–200 (1984). https://doi.org/10.1080/00207178408933267
16. Holmes, R.: A course on opimization and best approximation. Lecture Notes in Mathematics, vol. 257. Springer, Berlin (1972)
17. Howard, R.: Dynamic Probabilistic Systems. Wiley, New York (1971)
18. Kalman, R., et al.: Topics in Mathematical System Theory. McGraw-Hill Book Company, New York (1969)
19. Ku, R., Athans, M.: On the adaptive control of linear systems using the open-loop feedback optimal approach. IEEE Tran. Autom. Control **18**, 489–493 (1973)
20. Luenberger, D.: Optimization by Vector Space Methods. Wiley, New York (1969)
21. Marti, K.: Convex approximations of stochastic optimization problems. Methods Oper. Res. **20**, 66–76 (1975)
22. Marti, K.: Approximationen stochastischer Optimierungsprobleme. Hain Konigstein/Ts (1979)
23. Marti, K.: Stochastic optimization methods in robust adaptive control of robots. In: Groetschel, M.E.A. (ed.) Online Optimization of Large Scale Systems, pp. 545–577. Springer, Berlin (2001)
24. Marti, K.: Adaptive Optimal Stochastic Trajectory Planning and Control (AOSTPC) for Robots, pp. 155–206. Springer, Berlin (2004)
25. Marti, K.: Stochastic nonlinear model predictive control (snmpc). In: 79th Annual Meeting of the International Association of Applied Mathematics and Mechanics (GAMM), Bremen 2008, PAMM, vol. 8, Issue 1, pp. 10775–10776. Wiley-VCH (2008)
26. Marti, K.: Stochastic Optimization Methods, 2nd edn. Springer, Berlin (2008). https://doi.org/10.1007/978-3-540-79458-5
27. Marti, K.: Continuous-time control under stochastic uncertainty. In: Cochran, J.E.A., (ed.) Wiley Encyclopedia of Operations Research and Management Science (EORMS). Wiley, Hoboken (2010). https://doi.org/10.1002/9780470400531.eorms0839
28. Marti, K.: Optimal control of dynamical systems and structures under stochastic uncertainty: stochastic optimal feedback control. Adv. Engin. Softw. (AES) **46**, 43–62 (2012). https://doi.org/10.1016/j.advengsoft.2010.09.008

29. Marti, K.: Stochastic optimal structural control: Stochastic optimal open-loop feedback control. Adv. Engin. Softw. **44**(1), 26–34 (2012). https://doi.org/10.1016/j.advengsoft.2011.05.040. CIVIL-COMP
30. Marti, K., Gröger, D.: Einführung in die lineare und nichtlineare Optimierung. Springer, Berlin (2000)
31. Marti, K., Stein, I.: Computing stochastic optimal feedback controls using an iterative solution of the hamiltonian system (2013). https://doi.org/10.4203/ccp.102.109
32. Natanson, L.: Theorie der Funktionen einer reellen Veränderlichen. Verlag Harri Deutsch, Zürich (1977)
33. Øksendal, B.: Stochastic Differential Equations. Universitext. Springer, Berlin (2000)
34. Richalet, J., et al.: Model predictive heuristic control: applications to industrial processes. Automatica **14**, 413–428 (1978). https://doi.org/10.1016/0005-1098(78)90001-8
35. Schacher, M.: Stochastisch optimale Regelung von Robotern. No. 1200 in Fortschritt-Berichte VDI, Reihe 8, Mess-, Steuerungs- und Regelungstechnik. VDI Verlag GmbH, Düsseldorf (2011)
36. Smirnov, W.: Lehrgang der Höheren Mathematik, Teil IV. Deutscher Verlag der Wissenschaft, Berlin (1966)
37. Soong, T.: Active Structural Control: Theory and Practice. Wiley, New York (1990)
38. Stengel, R.: Stochastic Optimal Control: Theory and Application. Wiley, New York (1986)
39. Tse, E., Athans, M.: Adaptive stochastic control for a class of linear systems. IEEE Trans. Autom. Control **17**(1), 38–52 (1972)
40. Walter, W.: Gewöhnliche Differentialgleichungen. Springer, Berlin (2000)

Chapter 4
Random Search Methods for Global Optimization—Basics

Abstract Random Search Methods for solving deterministic optimization problems, as arising in the deterministic substitute problems of stochastic optimization and stochastic optimal control problems, are considered in this chapter and Chaps. 5–7. Besides mathematical optimization techniques, one of the major methods for solving deterministic parameter optimization problems is random search methods (RSM), for the following reason: Solving optimization problems from engineering and economics, one meets often the following situation: One should find the global optimum, hence, most of the deterministic programming procedures, which are based on local improvements of the performance index $F(x)$, will fail: Concerning the objective function F one has a **black-box**—situation, i.e., there is only few a priori information about the structure of F, especially there is no knowledge about the direct functional relationship between the control or input vector $x \in D$ and its index of performance $F(x)$; hence—besides the more or less detailed a priori information about F—the only way of getting objective information about the structure of F is via evaluations of its values $F(x)$ by experiments or by means of a numerical procedure simulating the technical plant. After explaining the basic (RSM)-algorithm, conditions are presented guaranteeing the convergence, in some stochastic sense, of the search method to a global optimum. As an example, the random search method bis applied to discrete optimization problems. Since, especially toward the optimum, the speed of convergence may become rather low, possibilities for acceleration of (RSM) are considered. A basic method, which will be further developed in the next chapters, is to control the distribution of the search variates.

4.1 Introduction

Solving optimization problems from engineering, as, e.g., parameter—or process—optimization problems

$$\min F(x) \quad \text{s.t. } x \in D, \tag{4.1}$$

where D is a subset of $\mathbb{R}^n$, one meets often the following situation:

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9_4

(a) One should find the **global** optimum in (4.1), hence most of the deterministic programming procedures, which are based on local improvements of the performance index $F(x)$, will fail.
(b) Concerning the objective function F one has a **black-box**—situation, i.e., there is only few a priori information about the structure of F, especially there is no knowledge about the direct functional relationship between the control or input vector $x \in D$ and its index of performance $F(x)$; hence—besides the more or less detailed a priori information about F—the only way of getting objective information about the structure of F is via evaluations of its values $F(x)$ by experiments or by means of a numerical procedure simulating the technical plant.

Consequently, engineers use in these situations often a certain search procedure for finding an optimal vector x, see, e.g., Box' EVOP method in [2] and the random search methods as first proposed by Anderson [1], Brooks [3] and Karnopp [5]. Obviously, deterministic search methods can be considered as special stochastic ones.

In the basic random search routine considered in this section—allowing not only local improvements as in mathematical programming—a sequence of n-random vectors $X_0, X_1, \ldots, X_t, \ldots$ in D is constructed according to the following recurrence relation:

$$X_{t+1} := \begin{cases} z_{t+1}, z_{t+1} \in D \text{ and } F(z_{t+1}) < F(X_t) \\ X_t, \text{ if } z_{t+1} \notin D \text{ or } F(z_{t+1}) \geq F(X_t), \end{cases} \tag{4.2a}$$

$t = 0, 1, 2, \ldots$, where the starting point $X_0 := x_0$ is a realization x_0 of the random vector X_0 having the given distribution $P_{X_0} := \pi_{\text{start}}$ concentrated on the domain D_{start}. In many cases we have $D_{\text{start}} \subset D$. If the search process starts at a given, fixed point x_0, then $\pi_{\text{start}} = \varepsilon_{x_0}$, where ε_{x_0} denotes the one-point measure at the point x_0. Moreover, $z_1 = Z_1(\omega), z_2 = Z_2(\omega), \ldots$ are realizations of n-random vectors $Z_1, Z_2, \ldots$ such that

$$\begin{aligned} &P\big(Z_{t+1} \in B | X_0 = x_0, X_1 = x_1, \ldots, X_t = x_t, Z_1 = z_1, Z_2 = z_2, \ldots, Z_t = z_t\big) \\ &= P\big(Z_{t+1} \in B | X_0 = x_0, X_1 = x_1, \ldots, Xt = x_t\big) = \pi_t(x^t, B), \end{aligned} \tag{4.2b}$$

where $x^t := (x_0, x_1, \ldots, x_t)$ and $\pi_t(x^t, \cdot)$ is a given transition probability distribution, as, e.g., a joint normal distribution with mean x_t and covariance matrix $Q = Q_t$.

According to Definition (4.2a), given the states $X^t = x^t$, first an n-vector z_{t+1} is generated randomly according to the distribution $\pi_t(x^t, \cdot)$. Then, if z_{t+1} drops into the area of success $G_F(x_t)$, where

$$G_F(x) := \big\{y \in D : F(y) < F(x)\big\}, \tag{4.3}$$

we move to $X_{t+1} = z_{t+1}$, otherwise we stay at $X_{t+1} = X_t$. Thus, the whole search process X_t) stays within the union $D_0 := D_{\text{start}} \cup D$ of the domain of the starting

points and the feasible domain of the basic optimization problem (4.1). If the set of starting points $D_{\text{start}} \subset D$ is contained in the feasible domain D, then $D_0 = D$.

Moreover, we observe that if $X_{t+1} \in G_F(x_t)$, then also $X_s \in G_F(x_t)$ for all $s > t$. Furthermore, if $F^* = \inf\{F(x) : x \in D\}$ and, for given levels $\varepsilon > 0$, $M < 0$, resp., the set of $\varepsilon-$, $M-$optimal solutions of (4.1) is defined by

$$B_{\varepsilon,M} := \{y \in D : F(y) \leq F^* + \varepsilon, \text{ if } F^* \in \mathbb{R}, F(y) \leq M, \text{ if } F^* = -\infty, \text{ resp.}\}, \tag{4.4a}$$

then

$$X_s \in B_{\varepsilon,M} \Rightarrow X_{s+1} \in B_{\varepsilon,M}, \quad s = 0, 1, 2, \ldots. \tag{4.4b}$$

Hence,

$$P(X_s \in B_{\varepsilon,M}) \leq P(X_{s+1} \in B_{\varepsilon,M}), \quad s = 0, 1, 2, \ldots . \tag{4.4c}$$

In the following we assume that the objective function F of (4.1) is a measurable function on $\mathbb{R}^n$.

4.2 The Convergence of the Basic Random Search Procedure

For considering the convergence behavior of the search method (4.2a), we examine the probability

$$P(X_t \in B_{\varepsilon,M}), \quad t = 0, 1, 2, \ldots,$$

that the t-th iterate X_t is an ε-, $M-$optimal solution, resp., of (4.1), where $\varepsilon > 0$, $M < 0$ are given numbers. According to the considerations at the end of Sect. 4.1 these probabilities form a nondecreasing, convergent sequence, and due to (4.4b) we have that

$$X_t \notin B_{\varepsilon,M} \Leftrightarrow X_0 \notin B_{\varepsilon,M}, X_1 \notin B_{\varepsilon,M}, \ldots, X_t \notin B_{\varepsilon,M}, \tag{4.5a}$$

hence,

$$\begin{aligned} P(X_t \in B_{\varepsilon,M}) &= 1 - P(X_0 \notin B_{\varepsilon,M}, X_1 \notin B_{\varepsilon,M}, \ldots, X_t \notin B_{\varepsilon,M}) \\ &= 1 - \int\limits_{x_0 \notin B_{\varepsilon,M}} P(X_1 \notin B_{\varepsilon,M}, \ldots, X_t \notin B_{\varepsilon,M} | X_0 = x_0) \pi(dx_0). \end{aligned} \tag{4.5b}$$

Denoting by $K_t(x^t, \cdots)$ the conditional distribution of X_{t+1} given $X_0 = x_0$, $X_1 = x_1, \ldots, X_t = x_t$, we have

$$K_t(x^t, B) = \pi_t\big(x^t, B \cap G_F(x_t)\big) + \Big(1 - \pi_t\big(\big(x^t, G_F(x_t)\big)\big)\Big)\varepsilon_{x_t}(B), \tag{4.6a}$$

where ε_x is the one-point-measure at x. Thus, with $\overline{B}_{\varepsilon,M} := D_0 \backslash \overline{B}_{\varepsilon,M}$, we get

$$P(X_1 \notin B_{\varepsilon,M}, \ldots, X_t \notin B_{\varepsilon,M} | X_0 = x_0) = \int\limits_{x_1 \in \overline{B}_{\varepsilon,M}} K_0(x_0, dx_1) \ldots$$

$$\int\limits_{x_{t-1} \in \overline{B}_{\varepsilon,M}} K_{t-2}(x^{t-2}, dx_{t-1}) \cdot \int\limits_{x_t \in \overline{B}_{\varepsilon,M}} K_{t-1}(x^{t-1}, dx_t). \tag{4.6b}$$

Considering first the t-th integral in the above equation, we obtain

$$\begin{aligned} \int\limits_{x_t \in \overline{B}_{\varepsilon,M}} K_{t-1}(x^{t-1}, dx_t) &= K_{t-1}(x^{t-1}, D_0 \backslash B_{\varepsilon,M}) \\ &= K_{t-1}(x^{t-1}, D_0) - K_{t-1}(x^{t-1}, B_{\varepsilon,M}) \\ &= 1 - K_{t-1}(x^{t-1}, B_{\varepsilon,M}). \end{aligned} \tag{4.6c}$$

Having $x_{t-1} \notin B_{\varepsilon,M}$ we get $\varepsilon_{x_{t-1}}(B_{\varepsilon,M}) = 0$ and $B_{\varepsilon,M} \subset G_F(x_{t-1})$, see the definitions (4.3), (4.4a). Hence, (4.6a)–(4.6c) yield

$$\begin{aligned} &\int\limits_{x_t \in \overline{B}_{\varepsilon,M}} K_{t-1}(x^{t-1}, dx_t) \le 1 - \pi_{t-1}(x^{t-1}, B_{\varepsilon,M}) \\ &\le 1 - \inf \left\{ \pi_{t-1}(x^{t-1}, B_{\varepsilon,M}) : x_s \in D_0 \backslash B_{\varepsilon,M},\ 0 \le s \le t-1 \right\} \end{aligned} \tag{4.7}$$

for all $x_s \in D_0 \backslash B_{\varepsilon,M},\ s = 0.1, \ldots, t-1$.

Defining now $\alpha_t,\ t = 0, 1, ...,$ by

$$\alpha_t := \alpha_t(B_{\varepsilon,M}) = \inf \left\{ \pi_t(x^t, B_{\varepsilon,M}) : x_s \in D_0 \backslash B_{\varepsilon,M},\ 0 \le s \le t \right\}, \tag{4.8}$$

from (4.6a), (4.6b) we now obtain

$$P(X_1 \notin B_{\varepsilon,M}, \ldots, X_t \notin B_{\varepsilon,M} | X_0 = x_0) \le \prod_{s=0}^{t-1} (1 - \alpha_s(B_{\varepsilon,M})) \tag{4.9a}$$

Hence, by (4.5b) and (4.9a) it is

$$\begin{aligned} &P(X_t \in B_{\varepsilon,M}) \ge \\ &1 - \int\limits_{x_0 \notin B_{\varepsilon,M}} P(X_1 \notin B_{\varepsilon,M}, \ldots, X_t \notin B_{\varepsilon,M} | X_0 = x_0) \pi_{\text{start}}(dx_0) \\ &\ge 1 - (1 - \pi_{\text{start}}(B_{\varepsilon,M})) \prod_{s=0}^{t-1} (1 - \alpha_s(B_{\varepsilon,M})). \end{aligned} \tag{4.9b}$$

Since $\log u \leq u - 1$, for $u > 0$ we have

$$(1 - \pi_{\text{start}}(B_{\varepsilon,M})) \prod_{s=0}^{t-1} (1 - \alpha_t) \leq \exp\Big(- \pi_{\text{start}}(B_{\varepsilon,M}) - \sum_{s=0}^{t-1} \alpha_s \Big) \tag{4.9c}$$

and therefore also

$$P(X_t \in B_{\varepsilon,M}) \geq 1 - \exp\Big(- \pi_{\text{start}}(B_{\varepsilon,M}) - \sum_{s=0}^{t-1} \alpha_s \Big). \tag{4.9d}$$

Thus, from (4.9d) we get the following convergence result.

Theorem 4.1 *The search process* (4.2a) *has the following convergence properties:*

(a) If for an $\varepsilon > 0$, $M < 0$, *resp.,*

$$\sum_{s=0}^{\infty} \alpha_s(B_{\varepsilon,M}) = +\infty, \tag{4.10}$$

then $\lim_{t\to\infty} P(X_t \in B_{\varepsilon,M}) = 1$.

(b) Suppose that $F^* \in \mathbb{R}$ *and*

$$\lim_{n\to\infty} P(X_n \in B_\varepsilon) = 1 \text{ for every } \varepsilon > 0. \tag{4.11}$$

Then $\lim_{n\to\infty} F(X_n) = F^*$ *a.s. (with probability one),*

(c) Assume that $F^* \in \mathbb{R}$ *and* F *is continuous and that the level sets* D_ε *are nonempty and compact for each* $\varepsilon > 0$. *Then* $\lim_{t\to\infty} F(X_t) = F^*$ *implies that also* $\lim_{t\to\infty}$ *dist* $(X_t, D^*) = 0$, *where dist* (X_t, D^*) *denotes the distance between* X_t *and the set* D^* *of global minimum points of* (4.1).

Proof For a proof of assertions (b) and (c), see [8]. □

Note 4.1

(a) For the case that the distribution π_t of Z_{t+1} does not depend on the states x^t, preliminary versions of the decisive inequality (4.9a) may already be found in the early Random Search literature, see, e.g., [3].

(b) Comparing the above theorem with the 0-1-laws of probability theory we observe that this result is essentially a consequence from the Borel-Cantelli-type laws, see, e.g., [9, p. 400] and [4, pp. 1–6, 51–52].

Working with random search procedures, one observes that the rate of convergence—especially near to the optimum—may be very poor. Hence, in the following we consider modified random search procedures with an improved convergence behavior.

4.2.1 Discrete Optimization Problems

Consider now the case that D contains a finite number r of elements $d_i \in \mathbb{R}^n$, thus,

$$D = \{d_1, d_2, \ldots, d_r\}. \tag{4.12a}$$

Furthermore, assume that $D_{\text{start}} \subset D$, hence $D_0 = D$, and let

$$P\big(Z_{t+1} \in B | X_0, X_1, \ldots, X_t\big) = P\big(Z_{t+1} \in B | X_t\big). \tag{4.12b}$$

Hence, (Z_t) and (X_t) are discrete-time stochastic processes. Therefore (Z_t) is described by a transition matrix (π_{ij}^t) from $X_t = i$ to $Z_{t+1} = j$, and the iterates (X_t) are described by the transition matrix (p_{ij}^t) from $X_t = i$ to $X_{t+1} = j$, $t = 0, 1, \ldots$. For $X_t = d_i$ we have $G_F(d_i) = G_F(i) = \big\{j : F(d_j) < F(d_i)\big\}$. The relationship between (π_{ij}^t) and (p_{ij}^t) reads then:

$$p_{ij}^t = p_{ij}^{(t,t+1)} = \begin{cases} 0, & \text{if } j \notin G_F(i) \text{ and } j \neq i \\ 1 - \sum\limits_{l \in G_F(i)} \pi_{il}^t, & \text{if } j = i \\ \pi_{ij}^t, & \text{if } j \in G_F(i). \end{cases} \tag{4.12c}$$

Assuming now stationary search variables $Z_t(\omega)$, i.e., in case $\pi_{ij}^t = \pi_{ij}$ for all $t = 0, 1, \ldots$, then also (X_t) is stationary and by searching for stationary distributions of (p_{ij}) we get this result.

Theorem 4.2 *Let $\pi_{ij} > 0$ for all $i, j = 1, ..., r$ or suppose that $\sum\limits_{j \in G_F(i)} \pi_{ij} > 0$ for all $1 \leq i \leq r$ such that d_i is not a solution of the optimization problem (4.1). Then $\big(X_t(\omega)\big)$ converges with probability one to a solution of problem (4.1).*

Proof Without limitation we may assume here that $0 < \varepsilon < F_{max} - F^*$. According to (4.8) and the above assumptions, for the minimum probabilities $\alpha_t = \alpha_t(\varepsilon)$ we have

$$\begin{aligned} \alpha_t(\varepsilon) &= \inf\big\{\pi_t(x^t, B_\varepsilon) : x_s \in D \backslash B_\varepsilon, 0 \leq s \leq t\big\} = \inf\big\{\pi(x_t, B_\varepsilon) : x_t \in D \backslash B_\varepsilon\big\} \\ &= \inf\Big\{\sum_{d_j \in B_\varepsilon} \pi_{ij} : d_i \notin B_\varepsilon\Big\} =: \alpha_0 > 0, \end{aligned} \tag{4.12d}$$

provided that $\pi_{ij} > 0$ for all indices i, j. Since $\alpha_t(\varepsilon) = \alpha_0 > 0$ for all $t = 0, 1, ...$, and each ε, $0 < \varepsilon < F_{max} - F^*$, the assertion follows now from Theorem 4.1. Since in the present case there are a finite number elements of feasible points d_i, $i..., r$, and the sets $G_F(i)$, B_ε are contained in each other for corresponding values of ε, $F(d_i)$, resp., the proof for the second case follows then also from (4.12d). □

Example 4.1 For illustration we may assume—without limitation—that the elements $d_1, d_2, \ldots, d_r$ of the feasible domain D are arranged such that

$$F(d_1) < F(d_2) < \ldots < Fd_r). \tag{4.13a}$$

Hence, d_1 is the unique minimum point, and the remaining points d_j are arranged in strictly increasing order of the function values $F(d_j)$. With the stationary transition probabilities $\pi_{ij}^{(t,t+1)} = \pi_{ij}$ from X_t to Z_{t+1}, the stationary transition matrix $P^t = P := (p_{ij})$ from X_t to X_{t+1} reads

$$P = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ \pi_{21}^t & 1-\pi_{21}^t & 0 & \ldots & 0 \\ \pi_{31}^t & \pi_{32}^t & 1-(\pi_{31}^t+\pi_{32}^t) & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \pi_{r1}^t & \pi_{r2}^t & \pi_{r3}^t & \ldots & \pi_{rr}^t \end{pmatrix}. \tag{4.13b}$$

Corresponding to Theorem 4.2 we find that $q^T := (1, 0, \ldots, 0)$ is a left fixed point of P and $\lim_{t\to\infty} X_t = q$ with probability 1.

4.3 Adaptive Random Search Methods

In this section we describe a general method how to find search variables (Z_t) such that the convergence of (X_t) toward a solution of our basic problem (4.1) is accelerated. This can be achieved by an adaptive selection of the probability distribution of the search variates $Z_1, Z_2, \ldots$. In order to control the sequence (Z_t) we assume that the probability distribution

$$\pi_t(x_0, x_1, \ldots x_t, \cdot) = \pi_t(a_t, x_0, x_1, \ldots, x_t, \cdot) \tag{4.14a}$$

of Z_{t+1} depends on a control parameter vector $a_t \in A_t(x_0, x_1, \ldots, x_t)$, where $A_t \subset A$ is the set of admissible controls at time t and given state-history $x^t := (x_0, x_1, \ldots, x_t)$. Moreover A_t is assumed to be contained in a fixed set A. By

$$\delta = (\delta_t)_{t\geq 0},\ \delta_t : \mathbb{R}^{n(1+t)} \to A,\ t = 0, 1, \ldots \tag{4.14b}$$

we denote a decision rule, composed of the control functions or strategies $\delta_t, t = 0, 1, \ldots$, such that the control parameter vectors a_t are given by

$$a_t := \delta_t(x^t) \in A_t(x^t) \ \text{ for } x_s \in D, \quad 0 \leq s \leq t, \quad t = 0, 1, \ldots. \tag{4.14c}$$

The set Δ of admissible decision rules δ is defined then by

$$\Delta := \big\{\delta : \delta = (\delta_t)_{t\geq 0}, \delta_t(x_0, x_1, \ldots, x_t) \in A_t(x_0, x_1, \ldots, x_t) \\ \text{for } x_s \in D,\ 0 \leq s \leq t,\ t = 0, 1, \ldots\big\}. \tag{4.14d}$$

Note 4.2 Since the transition probabilities $\pi_t(a_t, x_0, x_1, \ldots, x_t, \cdot)$ depend on the controls a_t, the expectation operator $E = E^\delta$ depends on the decision rule δ.

Looking for an *optimal* decision rule δ^*, clearly we have to guarantee that the process (X_t) generated by δ^* converges actually to a solution of (4.1).

Note that the reachability property in Theorem 4.1 holds, e.g., if the decision rules satisfies the condition, see (4.8),

$$\sum_{t=0}^{\infty} \inf \left\{ \pi_t\big(\delta_t(x^t), x^t, B_\varepsilon^F\big) : x^t = (x_0, x_1, \ldots, x_t), x_s \in D \backslash B_\varepsilon, 0 \le s \le t \right\} = +\infty, \tag{4.15a}$$

where, cf. (4.4a)

$$B_\varepsilon := \big\{ y \in D : F(y) \le F^* + \varepsilon \big\}. \tag{4.15b}$$

In the stationary case $\pi_t(a_t, x^t, \cdot) = \pi(a_t, x_t, \cdot)$ and $\delta_t(x^t) = \delta(x_t), t = 0, 1, \ldots$ (4.14a) is reduced to the much simpler condition

$$\inf \left\{ \pi\big(\delta(x), x, B_\varepsilon^F\big) : x \in D \backslash B_\varepsilon \right\} > 0. \tag{4.15c}$$

Appropriate utility- or reward-criterion for the evaluation of the individual steps $(X_t) \to X_{t+1}$ of the search process (X_t) are, e.g.,

(a) Probability of success

$$u_t(x_t, x_{t+1}) = \begin{cases} 1, x_{t+1} \in G_F(x_t) \\ 0, \text{ otherwise,} \end{cases} \tag{4.16a}$$

hence $E^\delta\big(u_t(a_t, X_t, X_{t+1}) | X_t\big) = P\big(X_{t+1} \in G(X_t) | X_t\big)$ is the (conditional) probability of a success in the state S_t.

(b) Step length

$$u_t(X_t, X_{t+1}) = \begin{cases} \|X_{t+1} - X_t\|^p, & X_{t+1} \in G_F(X_t) \\ 0, & \text{otherwise,} \end{cases} \tag{4.16b}$$

where $p \ge 1$ is a fixed number. Here $E\big(u_t(a_t, X_t, X_{t+1} | X_t\big)$ is the average step length of X_{t+1} into the area of success $G(X_t)$.

A modification of the above case is

(c) Relative step length

$$u_t(X_t, X_{t+1}) = \begin{cases} \left(\dfrac{\|X_{t+1} - X_t\|}{\|X_t\|} \right)^p, & X_{t+1} \in G_F(X_t) \\ 0, & \text{otherwise.} \end{cases} \tag{4.16c}$$

Obviously, any linear combination of the above three criteria yields criterion. In the following we suppose

$$F^* = \inf\{F(x) : x \in D\} > -\infty.$$

Search procedures with an improved performance can be constructed now by maximizing [7] the expected (total) reward function

$$U_\infty(x_0, \delta) := E^\delta \sum_{s=0}^{\infty} \varrho^s u_s(\delta_s(x^s), x_s, x_{s+1}), \tag{4.17a}$$

with respect to the decision rule $\delta = (\delta_s)$ involving the control functions δ_s satisfying the constraints (4.14c). Here, ϱ, $0 < \varrho < 1$, denotes still a certain *discount factor*.

For the maximization of the expectation of $U_\infty(x_0, \delta)$ next to we consider the $(T-t)$-stage search processes (X_s) starting at time t and running then up to time $T > t$. Hence, with $X^s = (X_0, X_1, \ldots, X_s)$, $X_0 := x_0$, and $a_s = \delta(X^s)$, $s = t, t+1, \ldots, T-1$, let

$$\begin{aligned} &U_T(t, x^t; \delta_t, \ldots, \delta_{T-1}) \\ &:= E^\delta\Big(\sum_{s=t}^{T-1} \varrho^{(s-t)} u_s(\delta_s(X^s), X_s, X_{s+1}) | X_0 = x_0, X_1 = x_1, \ldots, X_t = x_t\Big) \end{aligned} \tag{4.17b}$$

denote the conditional expected reward of this $(T-t)-$stage process, given the time history $X_s = x_s$, $s = 0, 1, \ldots, t$, and the control functions $\delta_t, \ldots, \delta_{T-1}$.

Denote by $K_t(a_t, x^t, \cdot)$ the transition probabilities for $X_t \to X_{t+1}$

$$K_t(a_t, x^t, B) = P(X_{t+1} \in B | X^t = x^t) \tag{4.18a}$$

of the process (X_t) based on search variates (Z_t) controlled by control inputs a_t, $t = 0, 1, \ldots$, where $X^t = (X_0, X_1, \ldots, X_t)$ and B is any Borel subset of $\mathbb{R}^n$. According to the basic definition (4.2a), (4.2b) of X_t and (4.14a)–(4.14d) it holds

$$\begin{aligned} K_t(a_t, x^t, B) = K_t(a_t, x^t, B) &= \pi_t\big(a_t, x^t, B \cap G(x_t)\big) \\ &+\Big(1 - \pi_t\big(a_t, x^t, G(x_t)\big)\Big)\varepsilon_{x_t}(B), \end{aligned} \tag{4.18b}$$

cf. (4.6b), where $a_t = \delta_t(x^t)$, and ε_x denotes again the one-point measure at the point $x \in \mathbb{R}^n$.

Due to the above definitions, the reward functions $U_T(t, x^t; \delta_t, \ldots, \delta_{T-1})$, $t = 0, 1 \ldots, T-2, T-1$, see (4.17b), satisfy the recurrence rations

$$\begin{aligned} U_T(t, x^t; \delta_t, \ldots, \delta_{T-1}) &= \int \Big(u_t\big(\delta_t(x^t), x_t, x_{t+1}\big) \\ +\varrho U_T\big(t+1, (x^t, x_{t+1}); \delta_{t+1}, \ldots, \delta_{T-1}\big)\Big) K_t(\delta(x^t), dx_{t+1}) &= \bar{u}_t(\delta_t(x^t), x^t) \\ +\varrho \int U_T\big(t+1, (x^t, y); \delta_{t+1}, \ldots, \delta_{T-1}\big) K_t(\delta(x^t), x^t, dy), \end{aligned} \tag{4.19a}$$

where

$$\bar{u}_t(a_t, x^t) = E\big(u_t(a_t, X_t, X_{t+1})|X^t = x^t\big) = \int u_t(a_t, x_t, y)K_t(a_t, x^t, dy). \tag{4.19b}$$

With the set Δ of admissible decision rules δ, cf. (4.14d), the value function of the $(T-t)$-stage process $X_t, \ldots, X_T$ with given state-history x^t is now defined by

$$V_t^T(x^t) := \sup_{\delta_t,\ldots,\delta_{T-1}} \big\{U_T(t, x^t; \delta_t, \ldots, \delta_{T-1}) : \\ \delta_s(x^s) \in A_s(x^s), x_s \in D^{(s+1)}, t \le s \le T-1\big\}, \tag{4.20}$$

where $D^{(s+1)}$ denotes the $(s+1)$-fold Carthesian product of D.

As mentioned already above, the set of restrictions in (4.20) should also include a condition guaranteeing that the whole search process (X_t) controlled by the decision rule $\delta = (\delta_t)$ satisfies a reachability condition according to Theorem 4.1. However, in many practical problems this condition may be deleted since the optimal decision functions δ_t^* defined by the optimization problem (4.20) can be shown to generate a search process (X_t^*) fulfilling a sufficient reachability condition.

From (4.19a), (4.19b), for the value functions $V_t^T(x^t)$ we get then the following recurrence relation:

Theorem 4.3 *Let $V_T^T(x^T) = 0$ for all $x^T \in \mathbb{R}^{(T+1)n}$. If for all steps t under consideration the maximum is attained in (4.20), then the following backwards recurrence relation holds*

$$V_t^T(x^t) = \sup_{a\in A_t(x^t)} \int \Big(u_t(a, x_t, y) + V_{t+1}^T\big((x^t, y)\big)\Big)K_t(a, x^t, dy) \tag{4.21}$$
$$= \sup_{a\in A_t(x^t)} \big(\bar{u}_t(a, x^t) + \int V_{t+1}^T(x^t, y)K_t(a, x^t, dy)\big),$$

$t = T-1, T-2, \ldots, 1, 0$, *where* $a = \delta_t(x^t)$.

Proof Omitting for simplification the constraint set in (4.20), from (4.19a), (4.19b) we get

$$\begin{aligned} V_t^T(x^t) &:= \sup_{\delta_t,\ldots,\delta_{T-1}} U_T(t, x^t; \delta_t, \ldots, \delta_{T-1}) = \sup_{\delta_t} \sup_{\delta_{t+1},\ldots,\delta_{T-1}} U_T(t, x^t; \delta_t, \ldots, \delta_{T-1}) \\ &= \sup_{\delta_t} \sup_{\delta_{t+1},\ldots,\delta_{T-1}} \Big(\bar{u}_t(\delta_t(x^t), x^t) \\ &+\varrho \int U_T\big(t+1, (x^t, y); \delta_{t+1}, \ldots, \delta_{T-1}\big)K_t(\delta(x^t), x^t, dy)\Big) \\ &= \sup_{a_t\in A_t(x^t)} \Big(\bar{u}_t(a_t(, x^t) \\ &+ \varrho \sup_{\delta_{t+1},\ldots,\delta_{T-1}} \int U_T\big(t+1, (x^t, y); \delta_{t+1}, \ldots, \delta_{T-1}\big)K_t(\delta(x^t), x^t, dy)\Big). \end{aligned} \tag{4.22a}$$

Now, according to (4.17b) we have

$$U_T(t+1,(x^t,y);\delta_{t+1},\ldots,\delta_{T-1}) = E_{x^t,y}u_T(t+1,(x^t,y),\delta_{t+1}(x^t,y),$$
$$\delta_{t+2}(x^t,y,X_{t+2}),\ldots,\delta_{T-1}(x^t,y,X_{t+2},\ldots,X_{T-1})), \tag{4.22b}$$

where $E_{x^t,y}$ denotes the conditional expectation given $X^{t+1}=(x^t,y)$ and X_j are random vectors defined by (4.2a), (4.2b) and u_T is the sum in (4.17b). Taking now, cf. (4.22a), the integral in (4.22b) with respect to y and then the supremum with respect to $\delta_{t+1},\ldots,\delta_{T-1}$ under the constraints $\delta_s(x^s)\in A_s(x^s)$, $x^s\in D^{(s+1)}$, $t+1\le s\le T-1$, see (4.20), the question is whether the integral and the supremum can be interchanged. Assuming that the suprema in (4.22b) are attained at $a_s^*=\delta_s^*(x^s)$, $x^s\in D^{(s+1)}$, $t+1\le s\le T-1$, with the conditional expectation operator E_{x^t} with respect to $X^t=x^t$, from (4.22b) we get

$$\begin{aligned}
&E_{x^t}U_T\big(t+1,(x^t,y);\delta_{t+1}^*,\ldots,\delta_{T-1}^*\big)\\
&\le \sup_{\delta_{t+1},\ldots,\delta_{T-1}} E_{x^t}U_T\big(t+1,(x^t,y);\delta_{t+1},\ldots,\delta_{T-1}\big)\big)\\
&\le E_{x^t}\sup_{\delta_{t+1},\ldots,\delta_{T-1}} U_T\big(t+1,(x^t,y);\delta_{t+1},\ldots,\delta_{T-1}\big)\big)\\
&= E_{x^t}U_T\big(t+1,(x^t,y);\delta_{t+1}^*,\ldots,\delta_{T-1}^*\big)\big).
\end{aligned} \tag{4.22c}$$

Thus, (4.22c) yields

$$\begin{aligned}
&\sup_{\delta_{t+1},\ldots,\delta_{T-1}} E_{x^t}U_T\big(t+1,(x^t,y);\delta_{t+1},\ldots,\delta_{T-1}\big)\big)\\
&= E_{x^t}\sup_{\delta_{t+1},\ldots,\delta_{T-1}} U_T\big(t+1,(x^t,y);\delta_{t+1},\ldots,\delta_{T-1}\big)\big) = V_{t+1}^T(x^t,y).
\end{aligned} \tag{4.22d}$$

The assertion follows now from equation (4.22a) and (4.22d). □

Remark 4.1 According to the definition (4.18b) of $K_t(a_t,x^t,\cdot)$ we have

$$\begin{aligned}
\int V_{t+1}^T(x^t,y)K_t(a,x^t,dy) &= \int_{y\in G(x_t)} V_{t+1}^T(x^t,y)\pi_t(a,x^t,dy)\\
&\quad + V_{t+1}^T(x^t,x_t)\Big(1-\pi_t\big(a,x^t,G(x_t)\big)\Big).
\end{aligned} \tag{4.23a}$$

Furthermore, assuming $u_t(a,x,x)=0$ for all $t=0,1,\ldots$, and $x\in\mathbb{R}^n$, we have, cf. (4.19b),

$$\bar{u}_t(a,x^t) = \int_{y\in G(x_t)} u_t(a,x_t,y)\pi_t(a,x^t,dy). \tag{4.23b}$$

In the important Markovian case, i.e., if

$$\pi_t(a,x^t,\cdot)=\pi_t(a,x_t,\cdot) \text{ and } A_t(x^t)=A_t(x_t), \tag{4.23c}$$

the value function V_t^T depends only on x_t, see (4.14a)–(4.14d), (4.17b), (4.18a), (4.18b), and (4.21) has the form

$$V_t^T(x_t) = \sup_{a \in A_t(x_t)} \left(\bar{u}_t(a, x_t) + \int V_{t+1}^T(y) K_t(a, x_t, dy)\right). \tag{4.23d}$$

In the one-stage case $t = T - 1$ equation (4.21) has the simple form

$$V_{T-1}^T(x^{T-1}) = \sup\left\{\bar{u}_{T-1}(a, x^{T-1}) : a \in A_{T-1}(x^{T-1})\right\}. \tag{4.23e}$$

4.3.1 *Infinite-Stage Search Processes*

The decision process defined by (4.21) is called the **sequential stochastic decision process associated with the random search procedure** (4.2a), (4.2b). An important variant of this decision process results in the infinite-stage stationary Markovian case.

Let $\pi_t(a_t, x^t, \cdot) = \pi(a_t, x_t, \cdot), A_t(x^t) = A(x_t), u_t(a_t, x_t, x_{t+1}) = u(a_t, x_t, x_{t+1})$, $\delta_t(x^t) = \delta(x_t), t = 0, 1, \ldots$. Moreover, let $0 < \varrho < 1$ be a certain discount factor. According to Theorem 4.3, the value function $V_t^T = V_t^T(x)$ of the (T-t)-stage process depends only on the state $x_t = x$ and fulfills the recurrence relation:

$$V_t^T(x) = \sup_{a \in A(x)} \left(\bar{u}(a, x) + \int V_{t+1}^T(y) K(a, x, dy)\right), \tag{4.24a}$$

$t = T - 1, T - 2, \ldots, 1, 0$, where $a = \delta(x_t)$. Introducing the stage transformation $(T - t) \to t$, the transformed value function

$$W_t(x) := V_{T-t}^T(x), t = 0, 1, \ldots \tag{4.24b}$$

satisfies (insert $s := T - t$ and replace then again $s \to t$) the forward recurrence relations

$$W_t(x) = \sup_{a \in A(x)} \left(\bar{u}(a, x) + \varrho \int W_{t-1}(y) K(a, x, dy)\right), t = 0, 1, \ldots, \tag{4.24c}$$

where the functional equation (4.24c) holds for each integer T, and we have, cf. Theorem 4.3, $W_0(x) = 0$.

Under certain conditions the sequence $\left(W_t(x)\right)$ is convergent to the function $W^*(x)$ satisfying the asymptotic functional equation

$$W^*(x) = \sup_{a \in A(x)} \left(\bar{u}(a, x) + \varrho \int W^*(y) K(a, x, dy)\right). \tag{4.24d}$$

Moreover, an optimal decision rule δ^* is then given by $\delta^*(x) = a^* \in A(x)$, where a^* is a solution of the maximization problem in (4.24d).

4.4 Convex Problems

For simplicity, here we only consider here the minimization of a real-valued convex function $F : \mathbb{R} \to \mathbb{R}$ with respect to $D = \mathbb{R}$. Assuming the second derivative F'' exists and $F''(x) > 0$ for all $x \in \mathbb{R}$, the interval $G(x) = \{y \in \mathbb{R} : F(y) < F(x)\}$ may be approximated by the interval

$$H(x) = \left\{ y \in K : F'(x)(y - x) + \frac{F''(x)}{2}(y - x)^2 < 0 \right\}. \tag{4.25a}$$

It is easy to see that

$$H(x) = \left\{ y \in \mathbb{R} \,:\, x < y < x - 2\frac{F'(x)}{F''(x)} \right\}, \qquad \text{if } F'(x) < 0, \tag{4.25b}$$

$$H(x) = G(x) = \emptyset, \qquad \text{if } F'(x) = 0, \tag{4.25c}$$

$$H(x) = \left\{ y \in \mathbb{R} \,:\, x - 2\frac{F'(x)}{F''(x)} < y < x \right\}, \qquad \text{if } F'(x) > 0. \tag{4.25d}$$

For the conditional distribution $\pi(a, x, \cdot)$ of the search variables (Z_t) given $X_t = x$ we choose now a normal distribution with mean $\mu = x$ and variance $\sigma^2 = a^2$. Hence, in this case our decision parameter a is then the standard deviation σ. Furthermore, according to the above approximation of $G(x)$ by $H(x)$, we approximate the utility function $u(a, x, y)$ of Sect. 4.3, by

$$\tilde{u}(a, x, y) = \begin{cases} |y - x| \,, & y \in H(x) \\ 0 \quad , & \text{otherwise.} \end{cases} \tag{4.26a}$$

Obviously, the stochastic decision process associated with the random search procedure (4.2a), (4.2b) is stationary and $\bar{u}(a, x)$ may be approximated by

$$\begin{aligned} \bar{\tilde{u}}(a, x) &:= \int_{y \in H(x)} \tilde{u}(a, x, y)\pi(a, x, dy) \\ &= \frac{\sigma}{\sqrt{2\pi}} \left(1 - \exp\left(-\frac{1}{2} \left(\frac{2F'(x)}{\sigma F''(x)} \right)^2 \right) \right). \end{aligned} \tag{4.26b}$$

Starting from $\widetilde{W}_0(x) = W_0(x) = 0$, the approximate $\widetilde{W}_1(x) = \sup_{\sigma>0} \bar{\tilde{u}}(\sigma, x)$ to the value function $W_1(x)$, see (4.24c) and the approximative decision function $\tilde{\sigma}_1 = \tilde{\sigma}_1(x)$, defined by $\widetilde{W}_1(x) = \bar{\tilde{u}}(\tilde{\sigma}_1, x)$, are given by the following theorem.

Theorem 4.4 *Let g be the function $g(t) = \frac{1}{\sqrt{2\pi}t}\left(1 - \exp(-\frac{1}{2}t^2)\right)$, and let denote $t^* > 0$ the number where g attains its maximum g^*. Then,*

$$\widetilde{W}_1(x) = g^* \left|\frac{2F'(x)}{F''(x)}\right| \text{ and } \tilde{\sigma}_1(x) = \frac{1}{t^*}\left|\frac{2F'(x)}{F''(x)}\right|. \tag{4.27}$$

Proof Using the transformation $\sigma \to t := \frac{1}{\sigma}\left|\frac{2F'(x)}{F''(x)}\right|$, according to (4.26b) and the above definition of the function $g = g(t)$, we have $\bar{\tilde{u}}(\sigma, x) = \left|\frac{2F'(x)}{F''(x)}\right| g(t)$. This yields the assertion. □

Obviously, according to (4.26a), $\widetilde{W}_1(x) = g^*\left|\frac{2F'(x)}{F''(x)}\right|$ is an approximate to the average step length $s_1(x)$ of the first step $X_0 \to X_1$ of the search process (X_t). Comparing this result with Newton's method $x \to y = x - \alpha(x)\frac{F'(x)}{F''(x)}$, $\alpha(x) > 0$ for the minimization of F, we observe that in Newton's method the step length $s_N(x) = |y - x| = \alpha(x)\left|\frac{F'(x)}{F''(x)}\right|$ has—up to a normalizing factor—the same form as $s_1(x)$.

Similar results are obtained from comparisons of Theorem 4.4 with deterministic and stochastic gradient procedures.

In general, the computation of the further iterates $\widetilde{W}_t$ and $\tilde{\delta}_t$, $t = 2, 3, \ldots$ will be in general hardly carried out in practice, because of its difficulty and because $\tilde{\sigma}(x) = \tau(x)\left|\frac{2F'(x)}{F''(x)}\right|$ with a normalizing factor $\tau(x) > 0$ is a reasonable approximate to the optimal decision rule. This is also confirmed by numerical experiments. On the other hand, for the quadratic case

$$F(x) = x^2$$

we can obtain the exact results. In fact, then we have that $H(x) = G(x)$ and $W_1(x) = \widetilde{W}_1(x) = 2g^*|x|$ as also $\sigma_1(x) = \tilde{\sigma}_1(x) = \frac{2}{t^*}|x|$. For solving now the functional equation (4.24b) we work therefore with the assumptions

$$W^*(x) = C|x| \text{ and } \sigma^*(x) = c|x|, \tag{4.28}$$

where C, c are positive constants.

Theorem 4.5 *The optimal value W^* and the optimal decision rule $\delta^*(x) = \sigma^*(x)$ of the infinite-stage stationary stochastic decision process associated with the random search procedure for the minimization of $F(x) = x^2$ has the form (4.28), where $c \approx \dfrac{8\sqrt{\pi}}{4\sqrt{\pi} - \varrho\sqrt{2}}$ and $C \approx \dfrac{1}{\sqrt{2\pi}} - \dfrac{\varrho}{4\pi}$.*

Proof See [6]. □

Note 4.3 As was mentioned in Sect. 4.1, often an analytic expression for F is not known and only the function values $F(x)$ may be obtained. Hence the derivatives $F'(x)$, $F''(x)$ in the "optimal" decision rule $\tilde{\sigma}(x) = \tau(x) \left| \frac{2F'(x)}{F''(x)} \right|$ must be estimated from observations of F.

References

1. Anderson, R.: Recent advances in finding best operating conditions. J. Am. Stat. Assoc. **48**(264), 789–798 (1953). http://www.jstor.org/stable/2281072
2. Box, G.: Evolutionary operation: a method for increasing industrial productivity. J. R. Stat. Soc. Ser. C **6**(2), 81–101 (1957). https://doi.org/10.2307/2985505
3. Brooks, S.: A discussion of random methods for seeking maxima. Oper. Res. **6**(2), 244–251 (1958). https://doi.org/10.1287/opre.6.2.244
4. Iosifescu, M., Theodorescu, R.: Random Process and Learning. Springer, Berlin (1969)
5. Karnopp, D.C.: Random search techniques for optimization problems. Automatica **1**(2–3), 111–121 (1963). https://doi.org/10.1016/0005-1098(63)90018-9
6. Marti, K.: On accelerations of the convergence in random search methods. Methods Oper. Res. **37**, 391–406 (1980)
7. Neumann, K.: Dynamische Optimierung. Bibliographisches Institut, Mannheim (1969)
8. Rappl, G.: Konvergenzraten von Random-Search-Verfahren zur globalen Optimierung. Ph.D. thesis, UniBw München (1984)
9. Richter, H.: Wahrscheinlichkeitstheorie. Springer, Berlin (1966)

Chapter 5
Controlled Random Search Methods as a Stochastic Decision Process

Abstract As already discussed in the preceding chapter, in order to develop procedures for increasing the rate of convergence of the basic search method, the stochastic search procedure is equipped with a mechanism for controlling the conditional probability distributions of the search variates at the iteration points, generating the new trial points for improving the current iteration point. In an attendant control or stochastic decision process, the parameters of the search variables can be selected to maximize criteria for measuring the progress of the search, such as the probability of a step into the area of success, or the mean step length into the area of success at a certain iteration point. Due to the black-box situation concerning the objective function F, we have a stochastic control or decision process under uncertainty concerning the objective function. Based on a Bayesian approach, with the obtained information from the search algorithm, the conditional distribution of F, given the information obtained during the search, can be determined.

5.1 The Controlled (or Adaptive) Random Search Method

In order to increase the rate of convergence of the basic search method (4.2a), according to Sect. 4.3 we consider the following procedure, cf. [2, 3]. Based on the basic random search method (4.2a), by means of the definitions (I)–(III) we describe first an (infinite-stage) **sequential stochastic decision process associated to** (4.2a).

(I) We use next to the fact that the transition probabilities $\pi_t(x^t, \cdot)$ depend

$$\pi_t(x^t, \cdot) = \pi_t(a, x^t, \cdot)$$

usually on certain parameters $a = (a_j)_{j \varepsilon J} \in A$, as, e.g., on certain (mixed) moments of the random vector Z_{t+1}. Let

$$\hat{h}^t = (x_0, x_1, \ldots, x_t, z_1, \ldots, z_{t-1})$$

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9_5

be the process history of X_t, Z_t up to time t. The idea, developed first in [2, 3], is now to run the random search not with a fixed parameter a, but to use an "optimal" control

$$a = a_t^*(x^t) \text{ or } a = a_t^*(\hat{h}^t)$$

of the parameter a such that a certain criterion measuring the progress of the search, as, e.g., the probability of a search success or the mean step length into the area of success at each step $X_t \to X_{t+1}$ is as large as possible. In the following,

$$h^t = (x_0, x_1, \ldots, x_t, z_1, z_2, \ldots, z_t, a_0, a_1, \ldots, a_{t-1})$$

denotes the total process history up to time t.

(II) To each step $x_t \to X_{t+1}$ there is associated a conditional mean (search-) gain

$$E\big(u_t(a_t, x_t, X_{t+1})|h^t\big)\ E\big(u_t(a_t, x_t, Z_{t+1})|h^t\big). \tag{5.1a}$$

Working, e.g., with the probability of a search success resp. the mean improvement of F resp. the mean (relative) step length into the area of success, u_t is given by

$$\begin{aligned} u_t(a_t, x_t, z_{t+1}) &= 1, = F(x_t) - F(z_{t+1}), = \|x_t - z_{t+1}\| \\ &= \frac{\|x_t - z_{t+1}\|}{\|x_t\|}, \text{ resp., if } z_{t+1} \in G_F(x_t) \\ u_t &= 0 \text{ if} z_{t+1} \notin G_F(x_t). \end{aligned} \tag{5.1b}$$

Calculating the conditional mean again in (5.1a) we have to solve next to integrals of the type

$$J(x^t, F) = \int\limits_{\substack{F(z_{t+1})<F(x_t) \\ z_{t+1}\in D}} u_t(a_t, x_t, z_{t+1})\pi_t(a_t, x^t, dz_{t+1}). \tag{5.2}$$

However, because of the black-box-situation concerning the objective function F of (4.1), i.e., having available at stage t only the discrete F-values $F(x_0), F(z_1), \ldots, F(z_t)$ as also the given some a priori information on F, the inequality $F(z_{t+1}) < F(x_t)$ in (5.2) can not be evaluated in general. Consequently, the integral $J(x^t, F)$ can not be computed in general because of the missing knowledge about F.

(III) In order to cope with this uncertainty, we may proceed in the following two different ways:

III.1 Approximation of the area of success $G_F(x_t)$, see (4.3), by a set $\widetilde{G}_F(x_t)$ which may be described by the information available on F up to the current time t.

As an example we mention here the following **Random-Search-Newton-Method** [3]: Similar to the (deterministic) Newton method in optimization, at state x_t we approximate first the value $F(z_{t+1})$ by the second-order Taylor polynomial

$$\begin{aligned} &F(z_{t+1}) \\ \approx &F(x_t) + \nabla F(x_t)^T(z_{t+1} - x_t) + \frac{1}{2}(z_{t+1} - x_t)^T \nabla^2 F(x_t)(z_{t+1} - x_t) \end{aligned} \tag{5.3a}$$

at x_t, where the derivatives ∇F, $\nabla^2 F$ of F may be obtained approximately by a numerical differentiation procedure using the process history h^t. Then, $G_F(x_t)$ can be approximated by the set

$$\begin{aligned} \widetilde{G}_F(x_t; \hat{h}^t) := \big\{ &y \in D : \widetilde{\nabla F}(x_t; \hat{h}^t)^T(y - x_t) \\ &+ \frac{1}{2}(y - x_t)^T \widetilde{\nabla^2 F}(x_t; \hat{h}^t)(y - x_t) < 0 \big\}, \end{aligned} \tag{5.3b}$$

where $\widetilde{\nabla F}(x_t; \hat{h}^t)$, $\widetilde{\nabla^2 F}(x_t; \hat{h}^t)$ are approximations to the gradient, Hessian, resp., $\nabla F(x_t)$, $\nabla^2 F(x_t)$ of F at x_t based on the process history h^t. The conditional mean gain (5.1a), (5.1b) be approximated now by

$$\begin{aligned} &E\big(u_t(a_t, x_t, Z_{t+1})|h^t\big) \approx E\big(\tilde{u}_t(a_t, x_t, Z_{t+1})|\hat{h}^t\big) \\ := &\int\limits_{z_{t+1} \varepsilon \widetilde{G}_F(x_t; \hat{h}^t)} u_t(a_t, x_t, z_{t+1}) \pi_t(a_t, x^t, dz_{t+1}). \end{aligned} \tag{5.3c}$$

By (I), (II), (III.1) we have now an infinite-stage sequential stochastic decision process for the definition of an optimal parameter control $a_t^* = a_t^*(h^t)$ speeding up the random search according to the chosen searching-gain criteria u_t. Since in practice our aim is to speed up to some extent the convergence of the basic random search procedure (4.2a) for solving (4.1), the computational effort for finding an *optimal* decision rule a_t^*, $t = 0, 1, \ldots$, should remain in realistic bounds. Hence, for practical purposes we are not interested in the exact solution $a_t^* = a_t^*(h^t)$ of the associated decision process, but we are looking for a sub-optimal control a_t^* obtainable by a reasonable computational effort. Approximating therefore the infinite-stage decision process by the family of 1-stage decision processes, a_t^* may be defined by

$$a_t^* = a_t^*(\hat{h}^t) \in \arg \max_{a \in A_t(x_t)} E\big(\tilde{u}_t(a, x_t, Z_{t+1})|\hat{h}^t\big), \tag{5.4}$$

where $A_t = A_t(x_t)$ denotes still the set of parameters available at (t, x_t). Having by the application of a_t^* a local improvement of the convergence

behavior of the random search, we will show later on also the convergence of the controlled process X_t^* toward $B_{\varepsilon,M}$.
Now the second method for handling the uncertainty concerning F is described.

III.2 Because of the missing information about the objective function F, actually we have a **sequential stochastic decision process under uncertainty**. Hence, the conditional mean search gain (5.1a), (5.1b) must be defined by

$$E\big(u_t(a_t, x_t, X_{t+1})|h^t\big) = \int J(x^t, F)\mu_{h^t}(dF), \tag{5.5}$$

where $J(x^t, F)$ is given by (5.2), and μ_{h^t} is the **conditional distribution of the unknown F**, *given the process history* $\mathbf{h^t}$. For the proper definition of μ_{h^t} we need first a mathematical representation of the given a priori information about F, as, e.g., "*F is an unknown polynomial in n variables*" or "*$a(x) \le F(x) \le b(x)$ for all $x \in D$ with given functions $a(\cdot), b(\cdot)$*".
We use a **Bayesian model** for the unknown F: We assume that there is a measurable space (θ, A) of parameters θ and an a priori distribution μ^0 of the parameters θ on A such that the objective function F of (4.1) is a realization

$$F(x) = f(x, \theta^0), \quad x \in \mathbb{R}^n, \tag{5.6a}$$

a. of a stochastic function $y = f(x, \theta), x \in \mathbb{R}^n, \theta \in \Theta$, on $\mathbb{R}^n$, where θ^0 is the true, but unknown parameter. We assume that each realization $f(\cdot, \theta)$ of f is a measurable function on $\mathbb{R}^n$.

Note 5.1 We observe that similar models for unknown functions in engineering have been used also in [1, 4].

In the next Sect. 5.2 we will show that

$$\mu_{h^t} = \mu_{\hat{h}^t},$$

i.e., the posterior distribution μ_h^t of F depends only on $\hat{h}^t = (x_0, x_1, \ldots, x_t, z_1, \ldots, z_t)$, but not on $a_0, a_1, \ldots, a_{t-1}$. Hence, due to (5.5) it is

$$E\big(u_t(a_t, x_t, X_{t+1})|h^t\big) = \int J(x^t, F)\mu_{h^t}(dF)$$
$$= \int J(x^t, F)\mu_{\hat{h}^t}(dF) = E\big(u_t(a_t, x_t, X_{t+1})|\hat{h}^t\big). \tag{5.7a}$$

Approximating as in approach III.1 the associated infinite-stage stochastic decision process by the family of 1-stage decision processes, a (sub-) optimal control a_t^* may be defined by

$$a_t^* = a_t^*(\hat{h}^t) = a_t^*(x_t, \mu_{\hat{h}^t}) \in \arg\max_{a \in A_t(x_t)} E\big(u_t(a_t, x_t, X_{t+1})|\hat{h}^t\big). \tag{5.7b}$$

Hence in both cases (5.4), (5.7b), the π_t-parameter control a_t^* depends on the (X_t, Z_t)-history $\hat{h}^t$ only.

5.1.1 The Convergence of the Controlled Random Search Procedure

As mentioned at the end of III.1 we have now to consider the convergence of the process $X_0^*, X_1^*, \ldots$ controlled by $a_t^*, t = 0, 1, \ldots$, toward the set $B_{\varepsilon,M}$ of (ε, M)-optimal solutions of (4.1), cf. (4.4a).

For this controlled procedure, denoted by Z_t^*, X_t^*, we consider similar to Sect. 4.3 first the conditional distribution $\widetilde{K}_t(\hat{h}^t, \cdot)$ of the tuple (Z_{t+1}^*, X_{t+1}^*), given $Z^{*t} = z^t$, $X^{*t} = x^t$ as also given the unknown F. Denoting by $T_{F,x}$ the mapping

$$T_{F,x}(y) = \begin{cases} y, \text{ if } y \in G_F(x), \\ x, \text{ else} \end{cases}, \tag{5.8}$$

obviously we have that $X_{t+1} = T_{F,x_t}(Z_{t+1})$ and therefore

$$\begin{aligned} &\widetilde{K}_t(\hat{h}^t, A \times B) := P(Z_{t+1}^* \in A, X_{t+1}^* \in B | Z^{*t} = z^t, X^{*t} = x^t) \\ &= P\big(Z_{t+1}^* \in A, T_{F,x_t}(Z_{t+1}^*) \in B | Z^{*t} = z^t, X^{*t} = x^t\big) \\ &\quad = \int\limits_{\substack{z_{t+1} \in \Lambda \\ T_{F,x_t}(z_{t+1}) \in B}} \pi_t\big(a_t^*(\hat{h}^t), x^t, dz_{t+1}\big) \\ &= \int\limits_{\substack{z_{t+1} \in A \cap B \\ z_{t+1} \in G_F(x_t)}} \pi_t(a_t^*, x^t, dz_{t+1}) + \left(\int\limits_{\substack{z_{t+1} \in A \\ z_{t+1} \notin G_F(x_t)}} \pi_t(a_t^*, x^t, dz_{t+1}) \right) \varepsilon_{x_t}(B), \end{aligned} \tag{5.9}$$

where the last equality follows from $z_{t+1} \in G_F(x_t) \Rightarrow T_{F,x_t}(z_{t+1}) = z_{t+1}$ and $z_{t+1} \notin G_F(x_t) \Rightarrow T_{F,x_t}(z_{t+1}) = x_t$.

Corresponding to (4.5a), (4.5b), (4.6a), and (4.6b), for the unknown, but fixed objective function F it holds then that

$$P(X_t^* \in B_{\varepsilon,M}|F) = 1 - \big(1 - 1_{B_{\varepsilon,M}}(x_0)\big) P(X_1^* \notin B_{\varepsilon,M}, \ldots, X_t^* \notin B_{\varepsilon,M}|F). \tag{5.10}$$

Putting $\overline{B}_{\varepsilon,M} := D\backslash B_{\varepsilon,M}$, we get

$$
\begin{aligned}
&P(X_1^* \notin B_{\varepsilon,M}, \ldots, X_t^* \notin B_{\varepsilon,M}|F) = \\
&= P\big((Z_1^*, X_1^*) \in \mathbb{R}^n \times \overline{B}_{\varepsilon,M}, (Z_2^*, X_2^*) \in \mathbb{R}^n \times \overline{B}_{\varepsilon,M}, \ldots, (Z_t^*, X_t^*) \\
&\in \mathbb{R}^n \times \overline{B}_{\varepsilon,M}|F\big) \\
&= \int\limits_{\substack{z_1 \in \mathbb{R}^n \\ x_1 \in \overline{B}_{\varepsilon,M}}} \widetilde{K}_0(x_0, dz_1, dx_1) \ldots \int\limits_{\substack{z_{t-1} \in \mathbb{R}^n \\ x_{t-1} \in \overline{B}_{\varepsilon,M}}} \widetilde{K}_{t-2}(\hat{h}^{t-2}, dz_{t-1}, dx_{t-1}), \\
&\cdot \int\limits_{\substack{z_t \in \mathbb{R}^n \\ x_t \in \overline{B}_{\varepsilon,M}}} \widetilde{K}_{t-1}(\hat{h}^{t-1}, dz_t, dx_t). \qquad (5.11)
\end{aligned}
$$

Proceeding now as in (4.7)–(4.9d), we obtain next to by means of $x_{t-1} \notin B_{\varepsilon,M}$ and (5.8)

$$
\begin{aligned}
\int\limits_{\substack{z_t \in \mathbb{R}^n \\ x_t \in \overline{B}_{\varepsilon,M}}} \widetilde{K}_{t-1}(\hat{h}^{t-1}, dz_t, dx_t) &= \widetilde{K}_{t-1}\big(\hat{h}^{t-1}, \mathbb{R}^n \times (D\backslash B_{\varepsilon,M})\big) \\
&= 1 - \widetilde{K}_{t-1}(\hat{h}^{t-1}, \mathbb{R}^n \times B_{\varepsilon,M}) \\
= 1 - \int\limits_{T_{F,x_{t-1}}(z_t) \in B_{\varepsilon,M}} \pi_{t-1}(a_{t-1}^*, x^{t-1}, dz_t) &= 1 - \pi_{t-1}\big(a_{t-1}^*, x^{t-1}, B_{\varepsilon,M} \cap G_F(x_{t-1})\big) \\
&= 1 - \pi_{t-1}(a_{t-1}^*, x^{t-1}, B_{\varepsilon,M}), \qquad (5.12a)
\end{aligned}
$$

where the last two equalities hold because of $x_{t-1} \notin B_{\varepsilon,M}$. Denoting by U a subset of $\mathbb{R}^n$ containing all supports of the conditional distributions $\pi_t(a_t, x_t, \cdot)$, $t = 0, 1, \ldots$, corresponding to α_t in Sect. 4.2.1

$$
\begin{aligned}
\alpha_t^* &= \alpha_t^*(\varepsilon, M, F) \\
&= \inf\Big\{\pi_t\big(\alpha_t^*(\hat{h}^t), x^t, B_{\varepsilon,M}\big) : z_s \in U, x_s \in D\backslash B_{\varepsilon,M},\ 1 \le s \le t\Big\} \qquad (5.12b)
\end{aligned}
$$

is the minimal probability for finding an (ε, M)-optimal solution of (4.1) at stage $t = 1, 2, \ldots$. Because of

$$
\int\limits_{\substack{z_t \in \mathbb{R}^n \\ x_t \in B_{\varepsilon,M}}} \widetilde{K}_{t-1}(\hat{h}^{t-1}, dz_t, dx_t) \le 1 - \alpha_{t-1}^*, t = 1, 2, \ldots, \qquad (5.12c)
$$

from (5.11) we obtain with $\alpha_0^* = \pi_0\big(a_0^*(x_0), x_0, B_{\varepsilon,M}\big)$ that

$$P(X_1^* \notin B_{\varepsilon,M}, \ldots, X_t^* \notin B_{\varepsilon,M}|F)$$
$$\leq \left(1 - \pi_0\big(a_0^*(x_0), x_0, B_{\varepsilon,M}\big)\right) \sum_{s=1}^{t-1}(1-\alpha_s^*) = \prod_{s=0}^{t-1}(1-\alpha_s^*). \tag{5.13}$$

Thus, by (5.10) and (5.13) yield, cf. (4.9a)–(4.9d),

$$P(X_t^* \in B_{\varepsilon,M}|F) \geq 1 - \prod_{s=0}^{t-1}(1-\alpha_s^*) \geq 1 - \exp\left(-\sum_{s=0}^{t-1}\alpha_s^*\right). \tag{5.14}$$

Hence, we proved the following result, see Theorem 4.1.

Theorem 5.1 (Convergence of the controlled Random Search Method)
If $\sum_{s=1}^{\infty}\alpha_s^*(\varepsilon, M, F) = +\infty$, *then* $\lim_{t\to\infty} P(X_t^* \in B_{\varepsilon,M}|F) = 1$.

Example 5.1 Let $\pi_t(a_t, x^t, \cdot)$ be a normal distribution with mean x_t and covariance matrix $Q = (\sigma_i^2\delta_{ik})$, where

$$a_t = a_t(\hat{h}^t) = \big(\sigma_1(\hat{h}^1), \sigma_2(\hat{h}^t), \ldots, \sigma_n(\hat{h}^t)\big)^T$$

and $\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ if $i \neq j$. If we know about F at least that $B_{\varepsilon,M}$ is bounded and has non-zero measure, then from the above theorem we get immediately this consequence.

Corollary 5.1 *Let* $B_{\varepsilon,M}$ *be bounded and have non-zero Lebesgue-measure. If the variance control* $a_t^* = \big(\sigma_1^*(\hat{h}^t), \ldots \sigma_n^*(\hat{h}^t)\big)^T$ *fulfills* $\sigma_t^*(\hat{h}^t) \geq \sigma_0 > 0$ *for all* $t = 1, 2, \ldots$ *and all values of* $\hat{h}^t$, *then* $\sum_{s=1}^{\infty}\alpha_s^* = \infty$ *and therefore* $\lim_{t\to\infty} P(X_t^* \in B_{\varepsilon,M}|F) = 1$.

For practical purposes—besides the convergence of X_t^* toward an (ε, M)-optimal solution of (4.1)—of great importance is a stopping criterion for the searching procedure.

5.1.2 A Stopping Rule

A suitable criterion for terminating the search at stage $t = T$ would be

$$P(X_1^* \notin B_{\varepsilon,M}) \leq P, \tag{5.15}$$

i.e., the probability that X_T^* is not an (ε, M)-optimal solution of (4.1) is not greater than a prescribed small value p. By (5.11) and (5.13) for this probability we have the estimate

$$P(X_T^* \notin B_{\varepsilon,M}) = P(X_T^* \notin B_{\varepsilon,M}|F) \leq \prod_{s=0}^{T-1} \big(1 - \alpha_s^*(\varepsilon, M, F)\big). \tag{5.16}$$

If the minimal probabilities $\alpha_s^*(\varepsilon, M, F)$ have a lower bound $\tilde{\alpha} > 0$, then from (5.15) we get obviously this next result.

Lemma 5.1 *Let* $\alpha_t^*(\varepsilon, M, F) > \tilde{\alpha} > 0$ *for all* $t = 0, 1, 2, \ldots$.
Then, the stopping criterion (5.15) *is satisfied if*

$$T \geq \frac{\log p}{\log(1 - \tilde{\alpha})}. \tag{5.17}$$

Due to the uncertainty concerning F we can not work in general with the expression in (5.15) and (5.16) directly. However, replacing the unknown function F by its conditional distribution μ_{h^T-1} known at stage $T - 1$, we may approximate (5.15) by the stopping rules

$$\int P\big(X_T^* \notin B_{\varepsilon,M}(\Gamma)|F\big)\mu_{h^{T-1}}(dF) \leq p, \tag{5.18a}$$

$$\int \prod_{s=0}^{T-1} \big(1 - \alpha_s^*(\varepsilon, M, F)\big)\mu_{h^{T-1}}(dF) \leq p, \tag{5.18b}$$

5.2 Computation of the Conditional Distribution of F Given the Process History: Information Processing

According to the uncertainty-model for the unknown objective function F of (5.6a) described in Sect. 5.1, (III.2), we assume that F is a realization $F(x) = f(x, \theta^0)$ of a stochastic function $y = f(x, \theta)$, $\theta \in \Theta$. Moreover, μ^0 denotes the a priori distribution (on a σ-algebra $\mathfrak{A}$ on Θ) of the stochastic parameter θ, and by $f = f(x)$ we denote any realization $f(\cdot, \theta)$ of the stochastic function $y = f(x, \theta)$. Instead of $\mu^0(d\theta)$, $\mu_{h^t}(d\theta)$, we also write $\mu^0(df)$, $\mu_{h^t}(df)$, resp., interpreting then μ^0, μ_{h^t} as conditional probability distributions on a certain σ-algebra $\mathfrak{A}_{\mathcal{F}}$ on an appropriate space $\mathcal{F}$ of possible objective functions f defined (at least) on the admissible domain D, $D \subset \mathbb{R}^n$.

Given the process history $h^t = (x_0, x_1, \ldots, x_t, z_1, z_2, \ldots, z_t, a_0, a_1, \ldots, a_{t-1})$ obtained from the search process $(Z_t, X_t, t = 1, 2, ...,$ to find the minimum of the function $F = F(x)$, according to the definition of the random search process we have the following relations:

$$T_{F,x_s}(z_{s+1}) = x_{s+1}, t_0, 1, \ldots, t - 1, \tag{5.19}$$

where $T_{F,x_s}(z_{s+1})$ is defined by (4.3), (5.8).

At each stage $s = 0, 1, \ldots, t-1$, there are then three different possibilities:

(i) $z_{s+1} \notin D$ *(search failure I).*
In this case it is $z_{s+1} \notin G_F(x_s)$ and therefore

$$x_{s+1} = T_{F,x_s}(z_{s+1}) = x_s.$$

But in the same way, for each realization f of $f(x, \theta)$ because of $z_{s+1} \notin D$ we find that

$$T_{f,x_s}(z_{s+1}) = x_s.$$

Hence, because of $x_{s+1} = x_s$, in this case (i) the constraint

$$T_{f,x_s}(z_{s+1}) = x_{s+1}$$

is satisfied automatically and can therefore be omitted.

(ii) $z_{s+1} \in D$ and $F(z_{s+1}) \geq F(x_s)$ *(search failure II).*
In this case we again have that

$$x_{s+1} = T_{F,x_s}(z_{s+1}) = x_s.$$

Now for f we have the constraint

$$T_{f,x_s}(z_{s+1}) = x_{s+1} = x_s$$

which holds if and only if

$$f(z_{s+1}) \geq f(x_s).$$

Indeed, assume that $f(z_{s+1}) < f(x_s)$. But this implies that $z_{s+1} \neq x_s$ and $x_{s+1} = T_{f,x_s}(z_{s+1}) = z_{s+1} \neq x_s$ which is a contradiction to the constraint $T_{f,x_s}(z_{s+1}) = x_{s+1} = x_s$.

(iii) $z_{s+1} \in D$ and $F(z_{s+1}) < F(x_s)$ *(search success).*
Here it is

$$x_{s+1} = F_{F,x_s}(z_{s+1}) = z_{s+1} \neq x_s,$$

hence for f we have the constraint

$$T_{f,x_s}(z_{s+1}) = x_{s+1} = z_{s+1} \neq x_s$$

which is possible if and only if

$$f(z_{s+1}) < f(x_s).$$

Indeed, assume that $f(z_{s+1}) \geq f(x_s)$. This implies $T_{f,x_s}(z_{s+1}) = x_s$, which is a contradiction.

Based on the segment x_s, z_{s+1}, z_{s+1} of the time history $\hat{h}^t$ and the observed values of the true, but unknown function $F(x) = f(x, \theta^0)$, we find that the following constraints for the parameter θ of the unknown model function $f = f(x, \theta)$:

$$T_{f,x_s}(z_{s+1}) = x_{s+1} \text{ is equivalent to } \begin{cases} \text{no constraint, if } z_{s+1} \notin D \\ f(z_{s+1}) \geq f(x_s), \text{ if } z_{s+1} \in D \\ \text{and } F(z_{s+1}) \geq F(x_s) \\ f(z_{s+1}) < f(x_s), \text{ if } z_{s+1} \in D \\ \text{and } F(z_{s+1}) < F(x_s). \end{cases} \tag{5.20a}$$

Consequently, given the time history $\Theta(h^t)$, $\hat{h}^t$, the set $\Theta h^t = \Theta(\hat{h}^t)$ of admissible parameters θ up to stage t is defined recursively by

$$\Theta(h^0) = \Theta(h^0) := \Theta \tag{5.20b}$$

$$\Theta(\hat{h}^{t+1}) = \begin{cases} \Theta(\hat{h}^t), \text{ if } z_{t+1} \notin D \\ \left\{\theta \in \Theta(\hat{h}^t) : f(z_{t+1}, \theta) \geq f(x_t, \theta)\right\}, \text{ if } z_t \in D \\ \text{and } F(z_{t+1}) \geq F(x_t) \\ \left\{\theta \in \Theta(\hat{h}^t) : f(z_{t+1}, \theta) < f(x_t, \theta)\right\}, \text{ if } z_t \in D \\ \text{and } F(z_{t+1} < F(x_t). \end{cases} \tag{5.20c}$$

Note 5.2 Obviously we have

$$\theta^0 \in \Theta(h^t) \subset \Theta(h^{t-1}), t = 0, 1, \ldots.$$

From the above consideration we get the following result:

Theorem 5.2 (Representation of the admissible set of parameters)

(a) Given the time history h^t, the recursion for the admissible parameter domains $\Theta(\hat{h}^t)$ can be given by

$$\Theta(h^0) = \Theta(h^0) := \Theta, \tag{5.21a}$$

$$\Theta(\hat{h}^{t+1}) = \left\{\theta \in \Theta(\hat{h}^t) : T_{f(\cdot,\theta),x_t}(z_{t+1}) = x_{t+1}\right\}, \quad t = 0, 1, ... \tag{5.21b}$$

(b) Let then the index sets $I_1(\hat{h}^t)$, $I_2(\hat{h}^t)$ be defined by

$$I_1(\hat{h}^t) = \left\{s : 0 \leq s \leq t-1, \ z_{s+1} \in D \text{ and } F(z_{s+1}) \geq \Gamma(x_s)\right\}, \tag{5.21c}$$

$$I_2(\hat{h}^t) = \left\{s : 0 \leq s \leq t-1, \ z_{s+1} \in D \text{ and } F(z_{s+1}) < F(x_s)\right\}. \tag{5.21d}$$

By these definitions the set $\Theta(\hat{h}^t)$ of admissible parameters at stage t can be represented in the following form:

$$\Theta(\hat{h}^t) = \big\{\theta \in \Theta : \ f(z_{s+1}, \theta) \geq f(x_s, \theta) \text{ for } s \in I_1(\hat{h}^t), \quad f(z_{s+1}, \theta) < f(x_s, \theta) \text{ for } s \in I_2(\hat{h}^t)\big\}. \tag{5.21e}$$

Having the set $\Theta(\hat{h}^t)$ of admissible parameters θ of the model $f = f(x, \theta)$ for the analytically not given objective function F, given the process history $\hat{h}^t$, we get this result:

Theorem 5.3 (Conditional probability of F) *With the a priori distribution μ^0 of the model parameters θ of the unknown objective function F, the conditional probability of F, given the process history $\hat{h}^t$, reads*

$$\mu_{\hat{h}^t}(W) = \frac{\mu^0\big(W \cap \Theta(\hat{h}^t)\big)}{\mu^0\big(\Theta(\hat{h}^t)\big)}, \tag{5.22}$$

for each measurable set $W \subset \Theta$.

Example 5.2

$$f(x, \theta) = \sum_{i=1}^{r} f_i(x)\theta_i \tag{5.23a}$$

for F, where $f_1, f_2, \ldots, f_r$ are known functions and $\theta_i, i = 1, \ldots, r$, are unknown real parameters with an a priori distribution μ^0 on $(\Theta, \mathcal{A}) = (\mathbb{R}^r, \mathcal{B}^r)$, then $\Theta(\hat{h}^t)$ is described by a finite number ($\leq t$) of linear inequalities.

Here, the conditional distribution $\mu_{\hat{h}^t}$ of F, given the search history $\hat{h}^t$, is the restriction of the a priori distribution μ^0 to set $\Theta(\hat{h}^t)$. Thus, if μ^0 has a probability density $\phi_0 = \phi_0(\theta)$, then the probability density $\phi_{\hat{h}^t} = \phi_{\hat{h}^t}(\theta)$ of $\mu_{\hat{h}^t}$ reads

$$\phi_{\hat{h}^t}(\theta) = \mathbf{1}_{\Theta(\hat{h}^t)}(\theta)\frac{\phi_0(\theta)}{\mu^0(\Theta(\hat{h}^t))}. \tag{5.23b}$$

References

1. Kushner, A.: A versatile stochastic model of a function of unknown and time varying form. J. Math. Anal. Appl. **9**, 379–388 (1962)
2. Marti, K.: Diskretisierung stochastischer Programme unter Berücksichtigung der Problemstruktur. Z. Angew. Math. Mech. **59**, T105–T108 (1979)
3. Marti, K.: Adaptive Zufallssuche in der Optimierung. ZAMM **60**, 357–359 (1980)
4. Mockus, J.: On bayesian methods of optimisation. In: Dixon, L., Szegö, G. (eds.) Towards Global Optimisation, pp. 166–181. North-Holland Publishing Company, Amsterdam (1975)

Chapter 6
Applications to Random Search Methods with Joint Normal Search Variates

Abstract As an application of the previous description of a general method to accelerate random search algorithms, in the following we consider search variates Z_{n+1} at an iteration point $X_n = x_n$ having a joint normal conditional distribution with mean and covariance matrix $(\mu, \Lambda) = (\mu(x_n), \Lambda(x_n))$. The mean search gain for a step $X_n \to X_{n+1}$ is determined by means of the mean decrease of the objective function. For simplification, instead of the infinite-stage optimal decision process for the selection of the parameters of the joint normal distribution, only the optimal one-step, $X_n \to X_{n+1}$, gains are taken into account, where the convergence rate of the fixed parameter and the optimized search method is evaluated. Since the optimal parameters of the normal distribution depend on the gradient and Hesse matrix of the objective function F, in a numerical realization of the optimal RSM, Quasi-Newton methods can be applied.

6.1 Introduction

Solving optimization problems arising from engineering and economics, as, e.g., parameter- or process-optimization problems,

$$\min F(x) \text{ s.t. } x \in D, \tag{6.1}$$

where D is a measurable subset of $\mathbb{R}^d$ and F is a measurable real function defined (at least) on D, one meets often the following situation:

(I) One should find the **global** minimum F^* and/or a **global** minimum point x^* of (6.1). Hence, most of the deterministic programming procedures, which are based on local improvements of the objective function $F(x)$, will fail.

(II) Concerning the objective function $F(x)$ one has a **black-box**-situation, i.e., there is only few a priori information about F especially there is no (complete) knowledge about the direct functional relationship between the control or input vector $x \in D$ and its function value $y = F(x)$. Hence, besides the limited a priori information about F, only by evaluating F numerically or by experiments at certain points $z_1, z_2, \ldots$ of $\mathbb{R}^d$ one gets further information on F.

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9_6

Consequently, engineers use in these situations usually a certain search procedure for finding the global minimum F and an optimal solution x^* of (6.1), see, e.g., Box' EVOP method [2] and the random search methods as first proposed by Anderson [1], Brooks [3] and Karnopp [5]. More recent descriptions of random search procedures were given by [6–8, 10, 12–14].

In the random search method considered here the sequence $X_0(\omega)$, $X_1(\omega)$, ..., $X_n(\omega)$, ... of random iterates is constructed according to the following recurrence schema:

$$X_{n+1}(\omega) = \begin{cases} z_{n+1}, & \text{if } z_{n+1} \in D \text{ and } F(z_{n+1}) < F\big(X_n(\omega)\big) \\ X_n(\omega), & \text{else}, \end{cases} \tag{6.2}$$

$n = 0, 1, \ldots$, where $x_0(\omega) = x_0 \in D$ is a given starting point in D and $z_1, z_2, \ldots, z_n, \ldots$ are realizations of a sequence of random d-vectors

$$Z_1(\omega), Z_2(\omega), \ldots, Z_n(\omega), \ldots$$

having conditional distributions

$$\begin{aligned} &P\big(Z_{n+1}(\omega) \in B | X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n, Z_1 = z_1, \ldots, Z_n = z_n\big) \\ &\quad = P\big(Z_{n+1}(\omega) \in B | X_n = x_n\big) = \pi_n(x_n, B) \end{aligned} \tag{6.3a}$$

for each Borel subset B of $\mathbb{R}^d$. Here, $\pi_n(x_n, \cdot)$, $n = 0, 1, ...$, is a sequence of transition probability measures to be selected by the user of the search procedure. In many concrete cases Z_{n+1} has a $d-$dimensional normal distribution with mean vector μ_n and covariance matrix Λ_n, i.e.,

$$\pi_n(x_n, \cdot) = N(\mu_n, \Lambda_n), \quad n = 0, 1, \ldots, \tag{6.3b}$$

where $\mu_n = \mu_n(x_n)$ and $\Lambda_n = \Lambda_n(x_n)$ are certain functions of the last state (n, x_n).

Let the area of success $G_F(x)$ at a point $x \in \mathbb{R}^d$ be defined by

$$G_F(x) = \big\{y \in D : F(y) < F(x)\big\}. \tag{6.4}$$

At an iteration point x_n by the random search procedure (6.2) a d-vector z_{n+1} is generated randomly according to the transition probability distribution $\pi_n(x_n, \cdot)$ and from x_n we move to $x_{n+1} = z_{n+1}$ provided that $z_{n+1} \le G_F(x_n)$. Otherwise we stay at $x_{n+1} = x_n$ and generate a new random point z_{n+2} according to the distribution $\pi_{n+1}(x_{n+1}, \cdot) = \pi_{n+1}(x_n, \cdot)$.

We observe that $X_{n+1} \in G_F(x_n)$ implies $X_t \in G_F(x_n)$ for all $t > n$. Let the set D_ε of ε-optimal solutions of our global minimization problem (6.1) be defined by

$$D_\varepsilon = \big\{y \le D : F(y) \le \Gamma^* + c\big\}, \tag{6.5}$$

where $c > 0$ and F^* is given by

$$F^* = \inf\{F(x) : x \in D\};$$

let $F^* > -\infty$. Note that $S_n \in D_\varepsilon$ implies $X_t \leq D_\varepsilon$ for all $t > n$. Hence,

$$P(X_n \in D_\varepsilon), \quad n = 1, 2, \ldots$$

is a non-increasing sequence for each fixed $\varepsilon > 0$.

6.2 Convergence of the Random Search Procedure (6.2)

Let $\alpha_n(D_\varepsilon)$ denote the minimal probability that at the n-th iteration step $X_n \to S_{n+1}$ we reach the set D_ε from any point $X_n = x_n$ outside this set, i.e.,

$$\alpha_n(D_\varepsilon) = \inf\{\pi_n(x_n, D_\varepsilon) : x_n \in D \backslash D_\varepsilon\}. \tag{6.6}$$

According to [6] we have

Theorem 6.1 *(a) If for an* $\varepsilon > 0$

$$\sum_{n=0}^{\infty} \alpha_n(D_\varepsilon) = +\infty, \tag{6.7}$$

then $\lim_{n\to\infty} P(X_n \in D_\varepsilon) = 1$ *for every* $\varepsilon > 0$.

(b) Suppose that

$$\lim_{n\to\infty} P(X_n \in D_\varepsilon) = 1 \text{ for every } \varepsilon > 0. \tag{6.8}$$

Then $\lim_{n\to\infty} F(X_n) = F^*$ *a.s. (with probability one) for every starting point* $x_0 \in n$.

(c) Assume that F *is continuous and that the level sets* D_ε *are nonempty and compact for each* $\varepsilon > 0$. *Then* $\lim_{n\to\infty} F(X_n) = F^*$ *implies that also* $\lim_{n\to\infty} dist\,(X_n, D^*) = 0$, *where dist* (X_n, D^*) *denotes the distance between* X_n *and the set* $D^* = D_0$ *of global minimum points* x^* *of (6.1).*

Example 6.1 If $\pi_n(x_n, \cdot) = \pi(\cdot)$ is a fixed probability measure, then $\lim_{n\to\infty} F(X_n) = F^*$ a.s. holds, provided that

$$\pi\Big(\{y \in D : F(y) \leq F^x + \varepsilon\}\Big) > 0 \quad \text{for each } \varepsilon > 0.$$

This is true, e.g., is D_ε has a non-zero Lebesque measure for all $\varepsilon > 0$ and π has a probability density ϕ with $\phi(x) > 0$ almost everywhere.

Note 6.1 Further convergence results of this type were given by [11, 13].

Knowing several (weak) conditions which guarantee the convergence a.s. of (X_n) to the global minimum F^*, to the set of global minimum points D^*, resp., one should also have some information concerning the rate of convergence of $F(X_n)$, (X_n) to F^*, D^*, respectively.

By [13] we have now the following result. Of course, as in the deterministic optimization, in order to prove theorems about the speed of convergence, the optimization problem (6.1) must fulfill some additional regularity conditions.

Theorem 6.2 *Suppose that $D^* \neq \emptyset$ and the transition probability measure $\pi(x_n, \cdot)$ is a d-dimensional normal distribution $N\big(\mu(x_n), \Lambda\big)$ with a fixed covariance matrix Λ.*

(a) Let D_ε be bonded for some $\varepsilon = \varepsilon_0 > 0$ and assume that F is convex in a certain neighborhood of D^*. Then

$$\lim_{n\to\infty} n^\gamma \big(F(X_n) - F^*\big) = 0 \quad a.s. \tag{6.9}$$

for each constant γ such that $0 < \gamma < \frac{1}{d}$ and every starting point x_0.

(b) Let D_ε be compact, $D^* = \{x^*\}$, where $s^* \in int(D)$ (= interior of D), and suppose that F is continuous and twice continuously differentiable in a certain neighborhood of x^*. Moreover, assume that F has a positive definite Hessian matrix at x^*. Then for each starting point $x_0 \in D$ it is

$$\lim_{n\to\infty} n^\gamma \big(F(X_n) - F^*\big) = 0 \qquad a.s. \text{ for each } 0 < \gamma < \frac{2}{d}, \tag{6.10a}$$

$$\lim_{n\to\infty} n^\gamma \|X_n - x^*\| = 0 \qquad a.s. \text{ for each } 0 < \gamma < \frac{1}{d}, \tag{6.10b}$$

$$\limsup_{n\to\infty} n^{\frac{2}{d}} E\big(F(X_n) - F^*\big) \leq \tau(x_0) < \infty, \tag{6.10c}$$

where $\tau(x_0)$ is a nonnegative finite constant depending on the starting point $x_0 \in D$ and E denotes the expectation operator.

(c) Under the same assumptions as in (b) we also have for each starting point $x_0 \in D$, $x_0 \neq x^*$,

$$\liminf_{n\to\infty} n^{\frac{2}{d}} L\big(F(X_n) - F^*\big) \geq h(x_0), \tag{6.11}$$

where $h(x_0)$ is a nonnegative constant depending on the starting point x_0. Furthermore for each $x_0 \in D$, $x_0 \neq x^*$, it is

$$\liminf_{n\to\infty} n^\gamma \|X_n - x^*\| = +\infty \quad \text{for each } \gamma > \frac{2}{d}. \tag{6.12}$$

Note 6.2 (a) Theorem 6.2 holds also for many non-normal classes of transition probability measures $\pi_n(x_n, \cdot)$), see [13].

(b) It turns out that under the assumptions of Theorem 6.2b the speed of convergence of (6.2) to the global minimum of (6.1) is **exactly** given by

$$E\big(F(X_n) - F^* = O(n^{-\frac{2}{d}}). \tag{6.13}$$

(c) The above convergence rates reflect the fact that in practice one observes that the speed of convergence may be very poor—especially near to the optimum of (6.1).

Hence, using random search procedures, a main problem is the control of the basic random search algorithm (6.2) such that the speed of convergence of (X_n) of F^*, D^*, resp., is increased.

6.3 Controlled Random Search Methods

A general procedure how to speed up the search routine (6.2) is described in [6–8]. The idea is to **associated with the random search routine** (6.2) **a sequential stochastic decision process** defined by the following items (I)–(III):

(I) We observe that the conditional probability distribution $\pi_n(x_n, \cdot)$ of Z_{n+1} given $X_n = x_n$ depends in general on a certain (vector valued) parameter a, i.e.,

$$\pi_n(x_n, \cdot) = \pi_n(a, x_n, \cdot), \quad a \in A, \tag{6.14}$$

where A is the set of admissible parameters a. The method, developed first in [6–8], is now to run the algorithm (6.2) not with a fixed parameter a, but to use an *optimal* control

$$a = a_n^*(x_n) \tag{6.15}$$

of a such that a certain criterion—to be explained in (II)—is maximized. Exemplary, $\pi_n(x_n, \cdot)$ is assumed to be a d-dimensional normal distribution with mean μ_n and covariance matrix Λ_n. Hence, in this case we have

$$a = (\mu, \Lambda) \in A := M \times \mathcal{Q}, \tag{6.16}$$

where $M \subset \mathbb{R}^d$ and $\mathcal{Q}$ is the set containing all symmetric, positive definite $d \times d$ matrices and the zero matrix.

(II) To search step $X_n \to X_{n+1}$ there is associated a mean search gain

$$U_n(a_n, x_n) = E\big(u(x_n, X_{n+1})|X_n = x_n\big), \tag{6.17}$$

where the gain function $u(x_n, x_{n+1})$ is defined, e.g., by

$$u(x_n, x_{n+1}) = \begin{cases} 1, & \text{if } x_{n+1} \in G_F(x_n) \\ 0, & \text{else} \end{cases} \tag{6.18a}$$

$$u(x_n, x_{n+1}) = \begin{cases} F(x_n) - F(x_{n+1}), & \text{if } x_{n+1} \in G_F(x_n) \\ 0, & \text{else} \end{cases} \tag{6.18b}$$

$$u(x_n, x_{n+1}) = \begin{cases} \|x_n - x_{n+1}\|, & \text{if } x_{n+1} \in G_F(x_n) \\ 0, & \text{else} \end{cases} \tag{6.18c}$$

Hence, in the first case $U_n(a_n, x_n)$ is the probability of a search success, in the second case $U_n(a_n, x_n)$ is the mean improvement of the value of the objective function and in case (6.18c) $U_n(a_n, x_n)$ is the mean step length of a successful iteration step $X_n \to X_{n+1}$.

(III) Obviously, the convergence behavior of the random search process (S_n) can be improved now by maximizing the mean total search gain

$$U_\infty = U_\infty(a_0, a_1, \ldots) := E \sum_{n=0}^{\infty} \rho^n u(X_n, X_{n+1})$$

subject to the controls $a_n = a_n(x_n) \in A,\ n = 0, 1, \ldots$, where $\rho > 0$ is a certain discount factor. This maximization can be done in principle by the methods of stochastic dynamic programming, see, e.g., [9].

6.4 Computation of Optimal Controls

In order to weaken the computational complexity, the infinite-stage stochastic decision process defined in Sect. 6.3 is replaced by the sequence of 1-stage decision problems

$$\max\ U_n(a_n, x_n) \quad \text{s.t. } a_n \in A,$$

$n = 0, 1, 2, \ldots$. Hence the *optimal* control $a_n^* = a^*(x_n)$ is defined as a solution of

$$\max_{a \in A} \int_{y \in G_F(x)} u(x, y)\pi(a, x, dy). \tag{6.19}$$

In the following we consider the gain function (6.18b), i.e.,

$$u(x, y) = F(x) - F(y).$$

Since an exact analytical solution of (6.19) is not possible in general, we have to apply some approximations. Firstly, the area of success $G_F(x)$ is approximated according to

$$G_F(x) \approx \left(y \in \mathbb{R}^d : \nabla F(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 F(x)(y - x) < 0 \right), \tag{6.20}$$

where $\nabla F(x)$ denotes the gradient of F and $\nabla^2 F$ is the Hessian matrix of F at x. We assume that $\nabla^2 F(x)$ is regular and $\nabla F(x) \neq 0$. Defining then the vector $w \in \mathbb{R}^d$ by

$$y - x = w - \nabla^2 F(x)^{-1} \nabla F(x), \tag{6.21}$$

the quadratic inequality contained in (6.20) has the form

$$w^T \frac{\nabla^2 F(x)}{r} w < 1,$$

where $r > 0$ is defined by

$$r = \nabla F(x)^T \nabla^2 F(x)^{-1} \nabla F(x).$$

By the Cholesky-decomposition of $\dfrac{\nabla^2 \Gamma(x)}{r}$ we can determine a matrix Γ such that

$$\frac{\nabla^2 F(x)}{r} = \Gamma \Gamma^T. \tag{6.22}$$

Defining

$$v = \Gamma^T w, \tag{6.23}$$

the approximation (6.20) of $G_F(x)$ can be represented according to (6.21) and (6.22) by

$$G_F(x) \approx \left(x_N + \Gamma^{-1^T} v : \|v\| < 1 \right), \tag{6.24}$$

where $\| \cdot \|$ is the Euclidean norm and x_N is given by

$$x_N = x - \nabla^2 F(x)^{-1} \nabla F(x).$$

It is then easy to verify that by the same transformation

$$v = v(\omega) := \Gamma^T (y(\omega) - x_N)$$

the search gain $u(x, y) = \Gamma(x) - F(y)$ can be approximated by

$$u(x, y) \approx \frac{r}{2}\left(1 - \|v\|^2\right). \tag{6.25}$$

By means of (6.24) and (6.25) the objective function $U(a, x)$, $a = (\mu, \Lambda)$, of (6.19) can be approximated by

$$\widetilde{U}(a, Q) = \frac{r}{2} \int\limits_{\|v\|<1} \left(1 - \|v\|^2\right) f(q, Q, v) dv, \tag{6.26}$$

where the probability density $f = f(q, Q, v)$ of the transformation $v = v(\omega)$ of $y = y(\omega)$ is given by

$$f(q, Q, v) = \frac{1}{(2\pi)^{d/2} (\det Q)^{1/2}} \exp\left(-\frac{1}{2}(v - q)^T Q^{-1}(v - q)\right).$$

Here the d-vector q and the positive definite $d \times d$ matrix Q are given by

$$q = Ev(\omega) = \Gamma^T(\mu - x_n), \tag{6.27}$$
$$Q = \operatorname{cov}(v(\cdot)) = \Gamma^T \Lambda \Gamma. \tag{6.28}$$

By the 1-1-transformations (6.27) and (6.28), the maximization problem (6.19) can be approximated by

$$\max_{q \in K, Q \in \mathcal{Q}} \widetilde{U}(q, Q), \tag{6.29}$$

where K and $\mathcal{Q}$ are defined by

$$K = K(M) = \left\{\Gamma^T(\mu - x_N) : \mu \in M\right\},$$
$$\mathcal{Q} = \{0\} \cup \{\Lambda : \Lambda \text{ positive definite } d \times d \text{ matrix}\}$$

and M is a certain subset of $\mathbb{R}^d$.

By the preceding considerations we obtain the following result:

Theorem 6.3 *Let q^*, Q^* be an optimal solution of (6.29) and define μ^*, Λ^* by*

$$\mu^* = x_N + \Gamma^{T\,-1} q^*,$$
$$\Lambda^* = (\Gamma Q^{*-1} \Gamma^T)^{-1}. \tag{6.30}$$

Then the 1-stage optimal control $a^*(x) = (\mu^*, \Lambda^*)$ is given approximately by (6.30).

In order to determine q^* and Q^*, we suppose now that the feasible set M for the mean value μ is defined by

$$M = \left\{\mu \in \mathbb{R}^d : \gamma_1^2 r < (\mu - x_N)^T \nabla^2 F(x)(\mu - x_N) \le \gamma_2^2 r\right\}, \tag{6.31}$$

where $0 < \gamma_1 \le \gamma_2$ are arbitrary, but fixed constants. In this case $K = K(M)$ is given by

$$K = \left\{q \in \mathbb{R}^d : \gamma_1 \le \|q\| \le \gamma_2\right\},$$

where $\|\cdot\|$ denotes the Euclidean norm. Note that the important case $M = \{x\}$, i.e., $\mu = x$(= last iteration point), corresponds to the case $\gamma_1 = \gamma_2 = 1$.

Assume now that M is given by (6.31). Since each Q has the form $Q = T\Delta T^T$, where T is an orthogonal matrix and Δ is a diagonal matrix, the minimization problem (6.29) is equivalent to

$$\begin{aligned} &\max \quad \widetilde{U}(q, \Delta) \\ &\text{s.t.} \quad \gamma_1 \leq \|q\| \leq \gamma_2, \\ &\quad \Delta \in \mathcal{Q}, \Lambda \text{ diagonal.} \end{aligned} \tag{6.32}$$

By a further approximation, we find then that an optimal solution q^*, Q^* of (6.32) is given approximately by this equations

$$\begin{aligned} q^* &= k^* 1, 1 = (1, \ldots, 1)^T, k^* \in \mathbb{R} \\ Q^* &= c^* I, I = \text{ identity matrix}, c^* > 0. \end{aligned} \tag{6.33}$$

Now (6.30), (6.33) and (6.22) yield

$$\begin{aligned} \mu^* &= x_N + k^* \Gamma^{T^{-1}} 1, \\ \Lambda^* &= \left(\Gamma \frac{1}{c^*} \Gamma^T\right)^{-1} = c^* (\Gamma\Gamma^T)^{-1} = c^* \left(\frac{v^2(F(x)}{r}\right)^{-1} \\ &= c^* r \nabla^2 F(x)^{-1}. \end{aligned}$$

Hence, we have this result.

Theorem 6.4 *The 1-stage optimal control $a^*(x) = (\mu^*, \Lambda^*)$ of the random search procedure (6.2) is given approximately by*

$$\begin{aligned} \mu^* &= x_N + k^* \Gamma^{T^{-1}}, \\ \Lambda^* &= c^* \Big(\nabla F(x)^T \nabla^2 F(x)^{-1} \nabla F(x)\Big) \nabla^2 F(x)^{-1}, \end{aligned} \tag{6.34}$$

where $k^* \in \mathbb{R}$, $c^* > 0$ are certain fixed parameters.

6.5 Convergence Rates of Controlled Random Search Procedures

Assume that the random search procedure (6.2) has normal distributed search variates $Z_1(\omega), Z_2(\omega), \ldots, Z_n(\omega), \ldots$ controlled by means of the following control law

$$\begin{aligned} \mu^0(x) &= x \\ \Lambda^0(x) &= c\Big(\nabla F(x)^T \nabla^2 F(x)^{-1} \nabla F(x)\Big) \nabla^2 F(x)^{-1}, \end{aligned} \tag{6.35}$$

where $c > 0$ is a fixed parameter. For control (6.35) we obtain, see the later considerations, this result:

Theorem 6.5 *Suppose that D is a compact, convex subset of $\mathbb{R}^d$ and let x^* be the unique optimal solution of (6.1). Let $x^* \in int(D)$ (= interior of D) and assume that F is twice continuously differentiable in a certain neighborhood of x^*. Moreover, suppose that $\nabla^2 F$ is positive definite at x^*. Then there is a constant $\kappa > 1$ such that*

$$\kappa^n E\big(F(X_n) - F^*\big) \to 0 \quad as\ n \to \infty$$

and

$$\kappa^n \big(F(X_n) - F^*\big) \to 0 \quad as\ n \to \infty, \quad a.s. \tag{6.36}$$

for all starting points contained in a certain neighborhood of x^*.

Note 6.3 (a) Comparing Theorem 6.2 and Theorem 6.5, we find that—at least locally—the convergence rate of (6.2) is increased very much by applying a suitable control, as, e.g., the control (6.35).

(b) However, the high convergence rate (6.36) holds only if the starting point x_0 is sufficiently close to x^*, while the low convergence rate found in Theorem 6.2 holds for arbitrary starting points $x_0 \in D$.

Hence, the question arises whether by a certain combination of a controlled random search procedure we also can guarantee a linear convergence rate for all starting points $x_0 \in D$.

Given an increasing sequence N of integers

$$n_1 < n_2 < \ldots < n_k < n_{k+1} < \ldots,$$

let the controls $a_n - (\mu_n, \Lambda_n)$ of the normal distributed search variates $Z_{n+1}(\omega)$, $n = 0, 1, 2, \ldots$, be defined by

$$\mu_n = x_n$$

and

$$\Lambda_n = \begin{cases} \Lambda^0(x_n)\ , & \text{if } n \in N \\ R\ , & \text{if } n \notin N, \end{cases} \tag{6.37}$$

where $\Lambda^0(x)$ is defined by (6.35) and R is a fixed positive definite $d \times d$ matrix.

Hence, according to (6.37), the search procedure is controlled only at the times $n_1, n_2, \ldots$.

Now, we have this result.

Theorem 6.6 *Suppose that D is a compact, convex subset o $\mathbb{R}^d$ and let $x^* \in int(D)$ be the unique optimal solution of (6.1). Assume that F is twice continuously differentiable in a certain neighborhood of x^* and let $\nabla^2 F(x^*)$ be positive definite. Define then*

$$h_n = \max\{k : n_k < n\}.$$

Then for every starting point $x_0 \in D$ there is a constant $\beta > 1$ such that

$$\beta^{h_n} E\big(F(X_n) - \Gamma^*\big) \to 0 \quad as\ n \to \infty$$

and

$$\beta^{h_n}\big(F(X_n) - F^*\big) \to 0 \quad as\ n \to \infty, \quad a.s., \tag{6.38}$$

provided that $\limsup_{n\to\infty} \dfrac{h_n}{n} < 1$.

Note 6.4 (a) Hence, the linear convergence rate (6.38) can be obtained by a suitable control of the type (6.37) for each starting points $x_0 \in D$.

(b) If $n_k = \dfrac{k}{p}$ for some $p \in \mathbb{N}$, then $\beta^{h_n} = (\sqrt[p]{\beta})^n$.

6.6 Numerical Realizations of Optimal Control Laws

In order to realize the control laws obtained in (6.34), (6.35), (6.37), one has to compute the gradient $\nabla F(x)$ and the inverse Hessian Matrix $\nabla^2 F(x)^{-1}$ of F at x. However, since the derivatives ∇F and $\nabla^2 F$ of F are not given in analytical form in practice, the gradient and the Hessian matrix of Γ must be approximated by means of the information obtained about Γ during the search process. Hence, for an approximate computation of υF and $\nabla^2 F$ we may use the sequence of sample points, iteration points and function values

$$x_0, F(x_0), z_1, F(z_1), x_1, z_2, F(z_2), x_2, \ldots\ .$$

In order to define a recursive approximation procedure, for $n = 0, 1, 2, \ldots$ let denote

$$\begin{aligned} g_n &\quad \text{the approximation of } \nabla F(x_n), \\ B_n &\quad \text{the approximation of } \nabla^2 F(x_n), \\ H_n &\quad \text{the approximation of } \nabla^2 F(x_n)^{-1}. \end{aligned}$$

Proceeding recursively, we suppose at the n-th stage of the search process we know the approximations g_n, B_n and H_n of $\nabla F(x_n)$, $\nabla^2 F(x_n)$ and $\nabla^2 F(x_n)^{-1}$, respectively. Hence, we may compute—approximately—the control $a_n = (\mu_n, \Lambda_n)$ according to one of the formulas (6.34), (6.35) or (6.37) by replacing $\nabla F(x_n)$ and $\nabla^2 F(x_n)^{-1}$ by g_n, H_n, respectively. The search process (6.2) yields then the sample point z_{n+1}, its function value $F(z_{n+1})$ and the next iteration point x_{n+1}. Now we have to perform the update

$$g_n \to g_{n+1},\ B_n \to B_{n+1} \text{ and } H_n \to H_{n+1} \tag{6.39}$$

by using the information s_n, $F(x_n)$, z_{n+1}, $F(z_{n+1})$, x_{n+1} about F.

(a) **Search failure at** x_n
If $z_{n+1} \notin D$ or $F(z_{n+1}) \geq F(x_n)$, then $x_{n+1} = x_n$. Since in this case we stay at x_n, we may define the update (6.39) by

$$\begin{aligned} g_{n+1} &= g_n, \\ B_{n+1} &= B_n, \\ H_{n+1} &= H_n. \end{aligned}$$

(b) **Search success at** x_n
In this case it is $z_{n+1} \in D$ and $F(z_{n+1}) < F(x_n)$, hence $x_{n+1} = z_{n+1} \neq x_n$. By a quadratic approximation of F at x_{n+1} we find then

$$\begin{aligned} F(x_n) \approx\ & F(x_{n+1}) + \nabla F(x_{n+1})^T (x_n - x_{n+1}) \\ & + \tfrac{1}{2}(x_n - x_{n+1})^T \nabla^2 F(x_{n+1})(x_n - x_{n+1}) \end{aligned}$$

and therefore

$$\nabla F(x_{n+1}^T s_n - \frac{1}{2} s_n^T \nabla^2 F(x_{n+1}) s_n \approx \Delta F_n, \tag{6.40}$$

where

$$\begin{aligned} s_n &= x_{n+1} - x_n = z_{n+1} - x_n, \\ \Delta F_n &= F(x_{n+1}) - F(x_n) = F(z_{n+1}) - F(x_n). \end{aligned}$$

Now we have to define the new approximations g_{n+1} and B_{n+1} of $\nabla F(x_n n + 1)$ and $\nabla^2 F(x_{n+1})$, respectively.
Because of (6.40), in order to define the update (6.39), we demand next to the following

Modified Quasi-Newton Condition

$$g_{n+1}^T s_n - \frac{1}{2} s_n^T B_{n+1} s_n = \Delta F_n \tag{6.41}$$

or

$$g_{n+1}^T s_n - \frac{1}{2} s_n^T B_{n+1} s_n < 0. \tag{6.42}$$

Note 6.5 (i) In contrary to (6.41), the modified Quasi-Newton condition (6.42) uses only the information that the function value of F at x_{n+1} is less than that at x_n.

(ii) If $\Delta F_n = F(x_{n+1}) \cdot F(x_n) < 0$, then $-s_n = x_n - x_{n+1}$ is an ascent direction of F at x_{n+1}. Hence, since $\nabla F(x_{n+1})$ is the best ascent direction of F at x_{n+1}, $-s_n$ may be used to define the approximation g_{n+1} of $\nabla F(x_{n+1})$.

Since g_{n+1} is not completely determined by the modified Quasi-Newton condition (6.41) or (6.42), resp., there are still many possibilities to define the update formulas (6.39). Clearly, since B_n is an approximation to a symmetric matrix, we suppose that B_n is a symmetric matrix.

(A) **Additive rank-one-updates**
In order to select a particular tuple (g_{n+1}, B_{n+1}) we may require that (g_{n+1}, B_{n+1}) is an optimal solution $(\overline{g}, \overline{B})$ of the distance-minimization problem

$$\begin{aligned} &\min \quad d_1(\overline{B}, B) + d_2(\overline{g}, g) \\ &\text{s.t.} \quad \overline{g}^T s - \tfrac{1}{2} s^T \overline{B} s = \Delta F, \end{aligned} \tag{6.43}$$

where $B = B_n$, $g = g_n$, $\Delta F = \Delta F_n$ and d_1, d_2 are certain distance measures. We suppose here that d_1, d_2 are defined by

$$\begin{aligned} d_1(\overline{B}, B) &= \tfrac{1}{2} \sum_{i,j-1}^{d} (\overline{b}_{ij} - b_{ij})^2, \\ d_2(\overline{g}, g) &= \tfrac{1}{2} \sum_{j-1}^{d} (\overline{g}_j - g_j)^2, \end{aligned} \tag{6.44}$$

where $\overline{b}_{ij}, b_{ij}$ are the elements of $\overline{B}$ and B, resp., and $\overline{g}_j, g_j$ denote the components of $\overline{g}, g$, respectively.

Note 6.6 The minimization (6.43) is a generalization of the minimality principles characterizing some of the well-known Quasi-Newton update formulas, see, e.g., [4].

Solving (6.43), (6.44), we find that $\overline{g}, \overline{B}$ are given by

$$\overline{g} = g - \lambda s \tag{6.45}$$

$$\overline{B} = B + \tfrac{\lambda}{2} s s^T, \tag{6.46}$$

where the Lagrange multiplier λ is given by

$$\lambda = \frac{g^T s - \frac{1}{2} s^T B s - \Delta\Gamma}{s^T s \left(1 + \frac{1}{4} s^T s\right)}. \tag{6.47}$$

If the distance functions d_1, d_2 are changed, then other update formulas may be generated. If, e.g., d_2 is replaced by $d_2(\overline{g}, g) = \dfrac{1}{2}(\overline{g} - g) B^{-1} (\overline{g} - g)$, then $\overline{g} = g - \lambda B s$.

Supposing now that B is positive definite, it is known that the matrix $\overline{B}$ defined by (6.46) is positive definite if and only if

$$1 + \frac{\lambda}{2} s^T H s > 0, \tag{6.48}$$

where $H = B^{-1}$ is our approximation to the inverse Hessian matrix $\nabla^2 F(x)^{-1}$ of F at $x = x_n$. Hence, if $\overline{H} = \overline{B}^{-1}$ denotes the approximation of the inverse Hessian matrix of F at x_{n+1}, then by (6.46) and (6.48) the following update formula $H \to \overline{H}$ for the inverse Hessian matrix of Γ may be established:

$$\overline{H} = \begin{cases} H - \frac{1}{2}\frac{\lambda}{1+\frac{\lambda}{2}s^T Hs} Hss^T H, & \text{if (6.48) holds} \\ H, & \text{else,} \end{cases} \tag{6.49}$$

Updates in the Case of a Search Failure

If $z_{n+1} \notin D$ or $F(z_{n+1}) \geq F(x_n)$, then we stay at $x_{n+1} = x_n$ and we may define therefore $\overline{g} = g$, $\overline{B} = B$ and $\overline{H} = H$. However, also in the case of a search failure the tuple $\big(z_{n+1}, F(z_{n+1})\big)$ yields new information about F, provided only that $z_{n+1} \neq x_n$. Hence, replacing the modified Quasi-Newton condition (6.41) by

$$\overline{g}^T s + \frac{1}{2} s^T \overline{B} s = \Delta F,$$

where now $s = z_{n+1} - x_n$, $\Delta F = F(z_{n+1}) - F(x_n)$, we may derive by the above procedure also update formulas $g \to \overline{g}$, $B \to \overline{B}$, $H \to \overline{H}$ for defining improved approximation $\overline{g}$, $\overline{B}$, $\overline{H}$ of ∇F, ∇^2 and $\nabla^2 F^{-1}$, respectively, at $x_{n+1} = x_n$.

(B) **Multiplicative rank-one-updates**

By (6.45)–(6.49) we have given a first concrete procedure for the realization of the optimal control laws (6.34), (6.35) and (6.37), respectively. Indeed, having, e.g., the mean $\mu_n = x_n$ and the covariance matrix

$$\Lambda_n = c^*(g_n^T H_n g_n) H_n, \tag{6.50}$$

the random variable Z_{n+1} may be defined by

$$Z_{n+1} = \mu_n + \Gamma_n Z_{n+1}^0$$

where Z_{n+1}^0 is a normal distributed with mean zero and covariance matrix equal to the identity matrix, and Γ_n is a $d \times d$ matrix such that

$$\Gamma_n \Gamma_n^T = \Lambda_n. \tag{6.51}$$

Hence, at each iteration point x_n the (Cholesky-)decomposition (6.51) of Λ_n have to be computed.

In order to omit this time consuming step, we still ask whether update formulas $\Gamma_n \to \Gamma_{n+1}$ for the Cholesky-factors Γ_n may be obtained.

Since $H_n = B_n^{-1}$ we suppose that B_n may be represented by

$$T_n T_n^T = B_n.$$

Then Λ_n is given by

$$\Lambda_n = c^*\big((T_n^{-1}g_n)^T T_n^{-1}g_n\big)T_n^{-1^T}T_n^{-1}$$

and the factor Γ_n may be defined, cf. (6.50) by

$$\Gamma_n = \sqrt{c^*}\|T_n^{-1}g_n\|T_n^{-1}. \tag{6.52}$$

In order to define the update $T \to \overline{T}$, where $T = T_n$ and $\overline{T} = T_{n+1}$ with $T_{n+1}T_{n+1}^T = B_{n+1}$, we require that T is changed only in the direction of $s = x_{n+1} - x_n$. Hence, we assume that

$$\overline{T} = (I + \frac{\gamma - 1}{s^T s}, ss^T)T,$$

where γ is real parameter to be determined. Furthermore, the distance-minimization problem (6.43) is then replaced by

$$\begin{aligned} \min \quad & d_1(\overline{T}, T) + d_2(\overline{g}, g) \\ \text{s.t.} \quad & \overline{g}^T s = \tfrac{1}{2}s^T \overline{B}s = \Delta F, \end{aligned} \tag{6.53}$$

where now

$$\overline{B} = \left(I + \frac{\gamma - 1}{s^T s}ss^T\right) B \left(I + \frac{\gamma - 1}{s^T s}ss^T\right), \quad \text{with } B = TT^T. \tag{6.54}$$

If the distance functions d_1, d_2 are again defined corresponding to (6.44), then

$$d_1(\overline{T}, T) = \frac{1}{2}\left(\frac{\gamma - 1}{s's}\right)^2 \|ss^T T\|_E^2, \tag{6.55}$$

where $\|T\|_E$ denotes the Euclidian norm of T. Hence, by (6.53) a particular tuple $(\overline{g}, \gamma)$ is selected. Because of (6.54) and (6.55), the minimization problem (6.53) has the form

$$\begin{aligned} \min \quad & \tfrac{\|ss^T T\|_E^2}{2}\left(\tfrac{\gamma - 1}{s's}\right)^2 + \tfrac{1}{2}\|\overline{g} - g\|_E^2 \\ \text{s.t.} \quad & \overline{g}^T s - \tfrac{\gamma^2}{2}s^T Bs - \Delta F, \end{aligned} \tag{6.56}$$

hence, the tuple $(\overline{g}, \gamma)$ is projected onto the parabola in $\mathbb{R}^{d+1}$ defined by the constraint in (6.56).

Note 6.7

(a) Other update formulas may be gained by changing the objective function of (6.56).
(b) Also in the case of a search failure, a similar method updates formulas may be derived.

References

1. Anderson, R.: Recent advances in finding best operating conditions. J. Am. Stat. Assoc. **48**(264), 789–798 (1953). http://www.jstor.org/stable/2281072
2. Box, G.: Evolutionary operation: a method for increasing industrial productivity. J. R. Stat. Soc. Ser. C **6**(2), 81–101 (1957). https://doi.org/10.2307/2985505
3. Brooks, S.: A discussion of random methods for seeking maxima. Oper. Res. **6**(2), 244–251 (1958). https://doi.org/10.1287/opre.6.2.244
4. Dennis, J., Schnabel, R.: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall, Englewood Cliffs (1983)
5. Karnopp, D.C.: Random search techniques for optimization problems. Automatica **1**(2–3), 111–121 (1963). https://doi.org/10.1016/0005-1098(63)90018-9
6. Marti, K.: Random search in optimization problems as a stochastic decision process (adaptive random search). Methods Oper. Res. **36**, 223–234 (1979)
7. Marti, K.: Adaptive Zufallssuche in der Optimierung. ZAMM **60**, 357–359 (1980)
8. Marti, K.: On accelerations of the convergence in random search methods. Methods Oper. Res. **37**, 391–406 (1980)
9. Müller, P., Nollau, V.: Steuerung stochastischer Prozesse. Akademie-Verlag, Berlin (1984)
10. Müller, P.E.A.: Optimierung mittels stochastischer Suchverfahren. Wissenschaftliche Zeitschrift der TU Dresden **21**(1), 69–75 (1983)
11. Oppel, U.: Random search and evolution. In: Trans. Symposia in Applied Mathematics. Thessaloniki (1976)
12. Rappl, G.: Optimierung durch Zufallsirchungsverfahren. Simulation und Optimierung det. und stoch. dynam, Systeme (1980)
13. Rappl, G.: Konvergenzraten von Random-Search-Verfahren zur globalen Optimierung. Ph.D. thesis, UniBw München (1984)
14. Zielinski, R., Neumann, P.: Stochastische Verfahren zur Suche nach dem Minimum einer Funktion. Akademie-Verlag, Berlin (1983)

Chapter 7
Random Search Methods with Multiple Search Points

Abstract Similar to the multi-start procedures in mathematical programming, here we consider random search methods working with multiple search variates (points) at an iteration point. The probability of failure, success, resp., and their properties at an iteration point are then evaluated for conditional independent, i.i.d., resp, stochastic search points. Furthermore, reachability results are given, i.e., results on the probability to reach an ϵ-optimal point with increasing stage or time. Finally, an optimized search process is studied based on the search point with minimum function value among all successful search points at the current iteration point.

7.1 Standard RSM

Given the optimization problem

$$\min\ F(x) \text{ subject to } x \in D \tag{7.1}$$

with the objective function $F = F(x)$ on $\mathbb{R}^n$ and the feasible domain $D \subset \mathbb{R}^n$, the Random Search Method (RSM) for solving (7.1) generates a (minimizing) random sequence (X_t) by the algorithm

$$X_{t+1} := T_{F,x_t}(Z_t), \quad \text{for } X_t = x_t,\ t = 0, 1, 2, \ldots, \tag{7.2a}$$

where $x_0 \in D$ is a given starting point,

$$T_{F,x}(z) := \begin{cases} z, & \text{for } z \in G_F(x) \\ x, & \text{else} \end{cases} \tag{7.2b}$$

and

$$G_F(x) := \{z \in D :\ F(z) < F(x)\}\,. \tag{7.2c}$$

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9_7

The sequence (X_t) can also be represented by

$$X_{t+1} = Z_t 1_{G_F(X_t)}(Z_t) + X_t 1_{G_F(X_t)^{\complement}}(Z_t), \quad t = 0, 1, 2, \ldots, \tag{7.3}$$

where $1_M = 1_M(x)$, $1_{M^{\complement}} = 1_{M^{\complement}}(x)$, resp. denote the characteristic functions of the set $M \subset \mathbb{R}^n$ and its complement $M^{\complement}$.

7.2 Multiple RSM

In contrast to standard RSM, in multiple RSM for each iteration stage $t = 0, 1, 2, \ldots$, we have a number $r \in \mathbb{N}$ of stochastic search variables at point X_t:

$$Y_{t,1}, Y_{t,2}, \ldots, Y_{t,j}, \ldots, Y_{t,r}, \qquad t \geq 0. \tag{7.4}$$

The random search process (X_t) is then defined as follows:

$$X_{t+1} := \begin{cases} Y_{t,1} & \text{if } Y_{t,1} \in G_F(X_t), \\ Y_{t,2} & \text{if } Y_{t,1} \notin G_F(X_t), \qquad Y_{t,2} \in G_F(X_t), \\ \vdots & \\ Y_{t,i} & \text{if } Y_{t,j} \notin G_F(X_t) \text{ for } 1 \leq j \leq i-1, \ Y_{t,i} \in G_F(X_t), \\ \vdots & \\ Y_{t,r} & \text{if } Y_{t,j} \notin G_F(X_t) \text{ for } 1 \leq j \leq r-1, \ Y_{t,r} \in G_F(X_t), \\ X_t & \text{if } Y_{t,j} \notin G_F(X_t) \text{ for } 1 \leq j \leq r. \end{cases} \tag{7.5}$$

For each stage $t = 0, 1, 2, \ldots$ the algorithm (7.5) is based on the events

$$S_{t,i} := \left[Y_{t,j} \notin G_F(X_t), \ 1 \leq j \leq i-1, \ Y_{t,i} \in G_F(X_t) \right], \quad i = 1, 2, \ldots r, \tag{7.6a}$$

of a search success after $i-1$ search failures. Moreover,

$$S_{t,f} := \left[Y_{t,j} \notin G_F(X_t), \ j = 1, 2, \ldots r \right] \tag{7.6b}$$

denotes the complete failure event that no search variable drops into the domain of success $G_F(X_t)$ at stage t.

According to the definitions (7.6a), (7.6b), the following properties hold:

Lemma 7.1

(a) For each given stage t, $t \geq 0$, $S_{t,i}$, $i = 1, 2, \ldots, r$ *are disjoint events.*

(b) $S_t := \bigcup_{i=1}^{r} S_{t,i}$ *denotes the event of a search success at stage* t. (7.6c)

(c) The event being represented by the complement of S_t,

$$S_t^{\complement} = \bigcap_{i=i}^{r} S_{t,i}^{\complement} = S_{t,f} \tag{7.6d}$$

represents the complete search failure at stage t.

Proof Properties in (a) and (b) follow directly from the definitions (7.6a) and (7.6c). From (7.6c) we get, with $G_F = G_F(x)$

$$\begin{aligned}
S_t^{\complement} &= \bigcap_{i=1}^{r} S_{t,i}^{\complement} \\
&= \left[Y_{t,1} \notin G_F\right] \quad \cap \quad \left(\left[Y_{t,1} \notin G_F\right] \cap \left[Y_{t,2} \in G_F\right]\right)^{\complement} \\
&\qquad\qquad \cap \quad \left(\left[Y_{t,1} \notin G_F\right] \cap \left[Y_{t,2} \notin G_F\right] \cap \left[Y_{t,3} \in G_F\right]\right)^{\complement} \\
&\quad \cap \ \ldots \quad \cap \quad \left(\left[Y_{t,1} \notin G_F\right] \cap \cdots \cap \left[Y_{t,i-1} \notin G_F\right] \cap \left[Y_{t,i} \in G_F\right]\right)^{\complement} \\
&\quad \cap \ \ldots \quad \cap \quad \left(\left[Y_{t,1} \notin G_F\right] \cap \cdots \cap \left[Y_{t,r-1} \notin G_F\right] \cap \left[Y_{t,r} \in G_F\right]\right)^{\complement},
\end{aligned}$$

hence,

$$\begin{aligned}
S_t^{\complement} &= \left[Y_{t,1} \notin G_F\right] \cap \left(\left[Y_{t,1} \in G_F\right] \cup \left[Y_{t,2} \notin G_F\right]\right) \\
&\qquad \cap \left(\left[Y_{t,1} \in G_F\right] \cup \left[Y_{t,2} \in G_F\right] \cup \left[Y_{t,3} \notin G_F\right]\right) \\
&\qquad \cap \ldots \\
&= \left[Y_{t,1} \notin G_F, Y_{t,2} \notin G_F\right] \cap \left(\left[Y_{t,1} \in G_F\right] \cup \left[Y_{t,2} \in G_F\right] \cup \left[Y_{t,3} \notin G_F\right]\right) \\
&\qquad \cap \ldots \\
&= \left[Y_{t,1} \notin G_F, Y_{t,2} \notin G_F, Y_{t,3} \notin G_F\right] \cap \ldots .
\end{aligned}$$

Proceeding this way, we find (7.6d). □

7.3 Probability of Failure, Probability of Success

According to (7.6b) and Lemma 7.1, the probability of failure p_f at point $X_t = x_t$ is given by

$$p_f = p_f(t, X_t) := P\left(Y_{t,j} \notin G_F(x_t),\ j = 1, 2, \ldots, r\right). \tag{7.7a}$$

If, as assumed in many (practical) cases, the search variables $Y_{t,j}$, $j = 1, \ldots r$, are stochastically independent for given $X_t = x_t$, then

$$\begin{aligned} p_f(t, x_t) &= \prod_{j=1}^{r} P\left(Y_{t,j} \notin G_F(x_t)\right) \\ &= \prod_{j=1}^{r} \left(1 - P(Y_{t,j} \in G_F(x_t))\right). \end{aligned} \tag{7.7b}$$

Consequently, the probability of success at point $X_t = x_t$ reads

$$\begin{aligned} p_s(t, x_t) =& 1 - p_f(t, x_t) \\ =& 1 - P\left(Y_{t,j} \notin G_F(x_t),\ j = 1, 2, \ldots, r\right), \end{aligned} \tag{7.8a}$$

and for stochastically independent search variables $Y_{t,j}$, $j = 1, 2, \ldots, r$, at $Xt = x_t$ we have

$$p_s(t, x_t) = 1 - \prod_{j=1}^{r} \left(1 - P\left(Y_{t,j} \in G_F(x_t)\right)\right). \tag{7.8b}$$

For the important special case

$$P\left(Y_{t,j} \in G_F(x_t)\right) = \alpha_F(t, x_t),\ j = 1, 2, \ldots, r, \tag{7.9}$$

e.g., for i.i.d. search variables at $X_t = x_t$, with a probability $\alpha_F(x_t)$ independent of index j, according to (7.7b), (7.8b) and (7.9), we have the following result.

Lemma 7.2 *In case of independent search variables $Y_{t,j}$, $j = 1, 2, \ldots, r$, having probability of success $\alpha_F = \alpha_F(t, x_t)$ at point $X_t = x_t$, the probability functions p_f and p_s are given by*

$$p_f(t, x_t) = (1 - \alpha_F)^r \tag{7.10a}$$

$$p_s(t, x_t) = 1 - (1 - \alpha_F)^r, \tag{7.10b}$$

with $\alpha_F = \alpha_F(t, x_t)$

According to the above lemma, the probability functions p_f, p_s can be represented by means of the binomial formula. Hence, for p_s we get

$$\begin{aligned} p_s(t, x_t) &= \tilde{p}_s(r, \alpha_F) \\ &= \sum_{j=1}^{r} \binom{r}{j} (-1)^{j+1} \alpha_F^j = r\alpha_F - \binom{r}{2}\alpha_F^2 + \binom{r}{3}\alpha_F^3 \pm \cdots + (-1)^{r+1}\alpha^r. \end{aligned} \tag{7.11}$$

For a search success with probability 1 at an index $j = j_0$, special values of p_f, p_s result, cf. (7.7b), (7.8b).

Remark 7.1 Let $Y_{t,j},\ j = 1, 2, \ldots, r$ be stochastically independent.
If $P\left(Y_{t,j} \in G_F(x_t)\right) = 1$ for $j = j_0,\ 1 \le j_0 \le r$, then

$$p_f(t, x_t) = 0 \qquad \text{and} \qquad p_s(t, x_t) = 1.$$

7.3.1 Monotonicity of the Probability Functions p_f, p_s

We now examine the dependence of p_s, p_f on r, the number of the used search variables $Y_{t,j},\ j = 1, \ldots, r$, for a given stage t. According to (7.7a) for the probability of failure we get

$$\begin{aligned} p_f(r+1, t, x_t) =& P\left(\left[Y_{t,j} \notin G_F(x_t),\ j = 1, \ldots, r\right] \cap \left[Y_{t,r+1} \notin G_F(x_t)\right]\right) \\ \le& P\left(Y_{t,j} \notin G_F(x_t),\ j = 1, \ldots, r\right) = p_f(r, t, x_t). \end{aligned} \tag{7.12a}$$

For stochastically independent search variables $Y_{t,j},\ j = 1, 2, \ldots, r$, this decrease of $p_f = p_f(r, t, x_t)$ with respect to r can be seen directly from (7.7b). In this case we have

$$p_f(r+1, t, x_t) < p_f(r, t, x_t) \tag{7.12b}$$

if $p_f(r, t, x_t) > 0$ and $0 \le P\left(Y_{t,r+1} \notin G_F(x_t)\right) < 1$.
Consequently, from (7.12a), (7.12b) we have

$$p_f(r, t, x_t) \le (<)\ p_f(\rho, t, x_t), \quad r > \rho. \tag{7.12c}$$

Comparing the multiple search $\left(Y_{t,j},\ 1 \le j \le r\right)$ with the single search $Y_t = Y_{t,1}$, we get

$$P\left(Y_t \notin G_F(x_t)\right) \ge (>)\ p_f(r, t, x_t),\ r > 1. \tag{7.12d}$$

Hence, with increasing number of search variables $Y_{t,j},\ j = 1, \ldots, r$ the search failure is decreasing. According to (7.8a), (7.8b) for the probability of a search success $p_s = p_s(r, t, x_t)$ at stage t we have

$$p_s(r, t, x_t) \ge (>)\ p_s(\rho, t, x_t) \ge (>)\ P\left(Y_t \in G_F(x_t)\right), \quad r > \rho. \tag{7.12e}$$

Thus, the probability of success can be increased by this method.

7.3.2 Asymptotic Behavior in Case of i.i.d. Search Variables

Suppose here that $Y_{t,j},\ 1 \le j \le r$, are i.i.d. stochastic search variables such that (7.9) holds. According to Lemma 7.2, we then have

$$\begin{aligned} p_f(t, x_t) &= \tilde{p}_f(r, \alpha_F) = (1 - \alpha_F)^r \\ p_s(t, x_t) &= \tilde{p}_s(r, \alpha_F) = 1 - (1 - \alpha_F)^r, \end{aligned}$$

with $\alpha_F = \alpha_F(t, x_t) > 0$. In order to consider the asymptotic behavior of p_f, p_s for $r \to \infty$, we use the inequality

$$(1 - xy)^n \le 1 - x + \mathrm{e}^{-yn}, \quad 0 \le x, y \le 1, \quad n > 0. \tag{7.13}$$

see [1], Lemma 10.5.3, p. 320. Applying (7.13) to $\tilde{p}_F = p_F(r, \alpha)$, we get

$$\tilde{p}_f(t, x_t) = (1 - \alpha_F)^r \le \mathrm{e}^{-r\alpha_F}, \quad r = 1, 2, \ldots . \tag{7.14}$$

Inequality (7.14) yields the following result.

Lemma 7.3 *Suppose $\alpha_F = \alpha_F(t, x_t) > 0$. Then*

$$p_f(t, x_t) = \tilde{p}_f(r, \alpha_F) \to 0, r \to \infty \tag{7.15a}$$

$$p_s(t, x_t) = \tilde{p}_s(r, \alpha_F) \to 1, r \to \infty. \tag{7.15b}$$

7.3.3 Estimation of p_f and p_s in Case of Arbitrary Stochastically Independent Search Variables $Y_{t,j} = Y_j$

In general we have

$$p_f(r, x_t) = \prod_{j=1}^{r} (1 - \alpha_j), \quad \alpha_j := P\left(Y_j \in G_F\right).$$

In the following we assume that

$$\underline{\alpha} \le \alpha_j \le \overline{\alpha}, \quad j = 1, 2, \ldots, r \tag{7.16a}$$

with given, fixed probability bounds $\underline{\alpha}, \overline{\alpha},\ 0 \le \underline{\alpha} < \overline{\alpha} \le 1$ for all values of $\alpha_j,\ 1 \le j \le r$. The above inequalities for $\alpha_j,\ 1 \le j \le r$ then yield

$$0 \le 1 - \overline{\alpha} \le 1 - \alpha_j \le 1 - \underline{\alpha} \le 1, \quad 1 \le j \le r.$$

Multiplying these inequalities by nonnegative values we get

$$(1-\overline{\alpha})^r \le \prod_{j=1}^{r}(1-\alpha_j) \le (1-\underline{\alpha})^r \tag{7.16b}$$

and therefore

$$(1-\overline{\alpha})^r \le p_f(r,x_t) \le (1-\underline{\alpha})^r. \tag{7.16c}$$

Moreover, for $p_s = 1 - p_f$ we get

$$1-(1-\underline{\alpha})^r \le p_s \le 1-(1-\overline{\alpha})^r. \tag{7.16d}$$

7.3.3.1 Error Estimation

Considering $p_f = p_f(r, x_t)$, we have the errors

$$\begin{aligned} e(r,\overline{\alpha}) &:= p_f - (1-\overline{\alpha})^r \text{ (left error)} \\ e(r,\underline{\alpha}) &:= (1-\underline{\alpha})^r - p_f \text{ (right error)} \end{aligned}$$

from using the lower, upper, resp. , approximation of p_f.

Thus

$$e(r,x_t) := e(r,\overline{\alpha}) + e(r,\underline{\alpha}) = (1-\underline{\alpha})^r - (1-\overline{\alpha})^r \tag{7.17a}$$

denotes the maximum error for approximating p_f by a value of the interval $[(1-\overline{\alpha})^r, (1-\underline{\alpha})^r]$.

Estimation of the maximum error $e(r, x_t)$ by the mean value theorem

$$\begin{aligned} e(r,x_t) &= (1-\underline{\alpha})^r - (1-\overline{\alpha})^r \\ &= -r\left(1-\left(\overline{\alpha}+\vartheta(\underline{\alpha}-\overline{\alpha})\right)\right)^{r-1}(\underline{\alpha}-\overline{\alpha}) \\ &= r\left(1-\left(\overline{\alpha}+\vartheta(\underline{\alpha}-\overline{\alpha})\right)\right)^{r-1}(\overline{\alpha}-\underline{\alpha}) \\ &\le r(1-\underline{\alpha})^{r-1}(\overline{\alpha}-\underline{\alpha}), \qquad 0<\vartheta<1. \end{aligned} \tag{7.17b}$$

Using (7.13) in Sect. 7.3.2, we finally have

$$\begin{aligned} e(r,x_t) &= (1-\underline{\alpha})^r - (1-\overline{\alpha})^r \\ &\le r\mathrm{e}^{-(r-1)\underline{\alpha}}(\overline{\alpha}-\underline{\alpha}) \\ &= \frac{r(\overline{\alpha}-\underline{\alpha})}{\mathrm{e}^{-(r-1)\underline{\alpha}}}. \end{aligned} \tag{7.17c}$$

Since the exponential function increases much faster than any power of r, for $\underline{\alpha} > 0$ we have

$$e(r, x_t) = (1 - \underline{\alpha})^r - (1 - \overline{\alpha})^r \to 0, \qquad r \to \infty. \tag{7.17d}$$

7.4 Reachability Results Multiple RSM

According to Chap. 4, here we have to study the sequence of probabilities

$$P\left(X_t \in B_\epsilon\right), \quad t = 1, 2, \ldots \tag{7.18a}$$

for

$$B_\epsilon := \left\{x \in D \,:\, F(x) \le F^* + \epsilon\right\}, \qquad \epsilon > 0, \tag{7.18b}$$

where the iterates $X_t,\ t = 1, 2, \ldots$ are given by (7.5). We have

$$\begin{aligned} P(X_t \in B_\epsilon) &= 1 - P(X_t \notin B_\epsilon) \\ &= 1 - P(X_1 \notin B_\epsilon, X_2 \notin B_\epsilon, \ldots, B_t \notin B_\epsilon), \end{aligned} \tag{7.19a}$$

with

$$\begin{aligned} &P(X_1 \notin B_\epsilon, X_2 \notin B_\epsilon, \ldots, B_t \notin B_\epsilon) \\ &= \int\limits_{x_1 \notin B_\epsilon} K_0(x_0, dx_1) \ldots \int\limits_{x_{t-1} \notin B_\epsilon} K_{t-2}(x_{t-2}, dx_{t-1}) \int\limits_{x_t \notin B_\epsilon} K_{t-1}(x_{t-1}, dx_t), \end{aligned} \tag{7.19b}$$

where $K_s(x_s, \cdot)$ denotes the probability distribution of X_{s+1}, given $X_s = x_s$. The partial integrals on $[x_s \notin B\epsilon],\ s = 1, 2, \ldots, t$, are now estimated from above stepwise $s = t, t-1, \ldots, 1$. For $s = t$ we get

$$\begin{aligned} &\int\limits_{x_t \notin B_\epsilon} K_{t-1}(x_{t-1}, dx_t) = K_{t-1}(x_{t-1}, D \setminus B_\epsilon) \\ &= K_{t-1}(x_{t-1}, D) - K_{t-1}(x_{t-1}, B_\epsilon) = 1 - K_{t-1}(x_{t-1}, B_\epsilon), \quad x_{t-1} \notin B_\epsilon, \end{aligned} \tag{7.20}$$

since $\subset D,\ x_0 \in D$ and $X_s \in D,\ s = 1, 2, \ldots$.

Because of $x_{t-1} \notin B_\epsilon$, hence,

$$F(x) \le F^* + \epsilon < F(x_{t-1}) \quad \text{for all } x \in B_\epsilon,$$

we have

$$B_\epsilon \subset G_F(x_{t-1}). \tag{7.21a}$$

Consequently, for $x_{t-1} \notin B_\epsilon$ we get

$$\begin{aligned} K_{t-1}(x_{t-1}, B_\epsilon) &= P(X_t \in B_\epsilon \mid x_{t-1}) \\ &= P\left(X_t \in G_F(x_{t-1}), X_t \in B_\epsilon \mid x_{t-1}\right). \end{aligned} \tag{7.21b}$$

According to the definition of $(X_t)_{t\geq 1}$, since $X_t \in G_F(x_{t-1})$ in this case we get $X_t = Y_{t,j}$ for one index $1 \leq j \leq r$, and

$$K_{t-1}(x_{t-1}, B_\epsilon) = P(S_{B_\epsilon} \mid x_{t-1}), \quad x_{t-1} \notin B_\epsilon, \tag{7.21c}$$

where S_{B_ϵ}, the set of elementary events having search variables reaching the set B_ϵ of ϵ-optimal points, is given by

$$\begin{aligned} S_{B_\epsilon} = &[Y_{t-1,1} \in B_\epsilon] \cup [Y_{t-1,1} \notin B_\epsilon, Y_{t-1,2} \in B_\epsilon] \cup \ldots \\ &\cup [Y_{t-1,1} \notin B_\epsilon, \ldots, Y_{t-1,r-1} \notin B_\epsilon, Y_{t-1,r} \in B_\epsilon]. \end{aligned} \tag{7.21d}$$

Corresponding to the domain of success $G_F(x_s)$, in case of stochastically independent search variables $Y_{t-1,j},\ 1 \leq j \leq r$, we have, cf. (7.8a), (7.8b),

$$\begin{aligned} K_{t-1}(x_{t-1}, B_\epsilon) &= 1 - \prod_{j=1}^{r} \left(1 - P(Y_{t-1,j} \in B_\epsilon \mid x_{t-1})\right) \\ &= 1 - \prod_{j=1}^{r} \left(1 - \Pi_{t-1,j}(x_{t-1}, B_\epsilon)\right), \end{aligned} \tag{7.22a}$$

where $\Pi_{t-1,j}(x_{t-1}, \cdot)$ denotes the probability distribution of $Y_{t-1,j}$, given $X_{t-1} = x_{t-1}$.

Thus, for $x_{t-1} \notin B_\epsilon$ we have

$$\int\limits_{x_t \notin B_\epsilon} K_{t-1}(x_{t-1}, dx_t) = \prod_{j=1}^{r} \left(1 - \Pi_{t-1,j}(x_{t-1}, B_\epsilon)\right). \tag{7.22b}$$

For the probability that X_t does not reach the set B_ϵ, provided that the realization $X_{t-1} = x_{t-1}$ is also outside of B_ϵ, we get the upper bound

$$\int\limits_{x_t \notin B_\epsilon} K_{t-1}(x_{t-1}, dx_t) \leq \prod_{j=1}^{r} \left(1 - \inf\left\{\Pi_{t-1,j}(x_{t-1}, B_\epsilon) \,:\, x_{t-1} \notin B_\epsilon\right\}\right). \tag{7.23a}$$

In the special case

$$\Pi_{t-1,j}(x_{t-1}, B_\epsilon) = \Pi_{t-1}(x_{t-1}, B_\epsilon), \quad j = 1, \ldots, r, \tag{7.23b}$$

we get

$$\int\limits_{x_t \notin B_\epsilon} K_{t-1}(x_{t-1}, dx_t) \leq (1 - \inf\{\Pi_{t-1}(x_{t-1}, B_\epsilon) \,:\, x_{t-1} \notin B_\epsilon\})^r \,. \tag{7.23c}$$

After estimating the last integral, i.e., the probability that X_t does not reach B_ϵ, given $X_{t-1} = x_{t-1} \notin B_\epsilon$, we get

$$\begin{aligned} &P\,(X_1 \notin B_\epsilon, X_2 \notin B_\epsilon, \ldots, X_t \notin B_\epsilon) \\ &\leq \int\limits_{x_1 \notin B_\epsilon} K_0(x_0, dx_1) \;\; \ldots \int\limits_{x_{-1} \notin B_\epsilon} K_{t-2}(x_{t-2}, dx_{t-1}) \prod_{j=1}^{r} (1 - \alpha_{t-1,j}), \end{aligned} \tag{7.24a}$$

where

$$\alpha_{t-1,j} := \inf\left\{\Pi_{t-1,j}(x_{t-1}, B_\epsilon) \,:\, x_{t-1} \notin B_\epsilon\right\}. \tag{7.24b}$$

Proceeding this way, we obtain

$$P\,(X_1 \notin B_\epsilon, X_2 \notin B_\epsilon, \ldots, X_t \notin B_\epsilon) \leq \prod_{s=0}^{t-1} \prod_{j=1}^{r} (1 - \alpha_{s,j}) \tag{7.25a}$$

with

$$\alpha_{s,j} := \inf\left\{\Pi_{s,j}(x_s, B_\epsilon) \,:\, x_s \notin B_\epsilon\right\} \quad s = 0, 1, \ldots, t-1. \tag{7.25b}$$

Since $\ln u \leq u - 1$, for $u > 0$, we have

$$\begin{aligned} \ln\left(\prod_{s=0}^{t-1} \prod_{j=1}^{r} (1 - \alpha_{s,j})\right) &= \sum_{s=0}^{t-1} \sum_{j=1}^{r} \ln(1 - \alpha_{s,j}) \\ &\leq \sum_{s=0}^{t-1} \sum_{j=1}^{r} (-\alpha_{s,j}) = -\sum_{s=0}^{t-1} \sum_{j=1}^{r} \alpha_{s,j} \end{aligned} \tag{7.26a}$$

and therefore

$$\prod_{s=0}^{t-1}\prod_{j=1}^{r}(1-\alpha_{s,j}) \le \exp\left(-\sum_{s=0}^{t-1}\sum_{j=1}^{r}\alpha_{s,j}\right). \tag{7.26b}$$

With the above inequalities, for $P(X_t \in B_\epsilon)$ we now find

$$\begin{aligned} P(X_t \in B_\epsilon) &= 1 - P(X_1 \notin B_\epsilon, \ldots, X_t \notin B_\epsilon) \\ &\ge 1 - \prod_{s=0}^{t-1}\prod_{j=1}^{r}(1-\alpha_{s,j}) \ge 1 - \exp\left(-\sum_{s=0}^{t-1}\sum_{j=1}^{r}\alpha_{s,j}\right). \end{aligned} \tag{7.27}$$

Consequently, we now get the following result.

Theorem 7.1 *Suppose that*

$$(i)\ \sum_{s=0}^{t-1}\sum_{j=1}^{r}\alpha_{s,j} \to \infty, \quad t \to \infty, \text{ for a given integer } r \ge 1, \text{ or} \tag{7.28a}$$

$$(ii)\ \sum_{s=0}^{t-1}\sum_{j=1}^{r}\alpha_{s,j} = \sum_{j=1}^{r}\sum_{s=0}^{t-1}\alpha_{s,j} \to \infty, \quad r \to \infty, \text{ for a given stage } t \ge 1, \tag{7.28b}$$

then, with $X_t = X_{t,r}$ we have

$$P(X_{t,r} \in B_\epsilon) \to 1, \quad t \to \infty, r \to \infty, \text{ respectively,} \tag{7.28c}$$

In the special case

$$\alpha_{s,j} = \alpha_s, \quad j = 1, \ldots, r \tag{7.29a}$$

with a fixed probability $\alpha_s > 0$, we get

$$P(X_{t,r} \in B_\epsilon) \ge 1 - exp\left(-r\sum_{s=0}^{t-1}\alpha_s\right). \tag{7.29b}$$

In this case we have $P(X_{t,r} \in B_\epsilon) > 1 - \delta$ with a given $\delta > 0$, provided that

$$r\sum_{s=0}^{t-1}\alpha_s > \ln\frac{1}{\delta}. \tag{7.29c}$$

7.5 Optimal Search Point Among Multiple Search Variables

Among the search variables $Y_{t,j} = Y_{t,j}(\omega),\ j = 1, \ldots, r,\ \omega \in \Omega$, at time t, see (7.4), the best search variable is given at a point x_t by

$$Y_t^* = Y_{t,j^*} := \underset{Y_{t,j} \in G_F(x_t), 1 \le j \le r}{\operatorname{argmin}} F(Y_{t,j}). \tag{7.30}$$

If the minimum is attained at several indices j, $j^* = j^*(t, \omega)$ denotes the smallest index.

7.5.1 The Optimized Search Process

Corresponding to the search process (X_t) with multiple search variables $Y_{t,j},\ j = 1, \ldots, r$, see (7.4), (7.5), using the best search variable $Y_t^* = Y_{t,j^*},\ j^* = j^*(t, \omega)$, defined by (7.30), we now consider the optimized search process (X_t^*) with the realizations $X_t^*(\omega) = x_t^*$, defined, cf. Fig. 7.1, by

$$X_0^* := x_0^* = x_0 \tag{7.31a}$$

$$x_{t+1}^* := \begin{cases} Y_{t,j^*(t,\omega)}, & \text{if there exists at least one variable} \\ & Y_{t,j} \in G_f(x_t^*), \quad 1 \le j \le r \\ x_t^*, & \text{else.} \end{cases} \tag{7.31b}$$

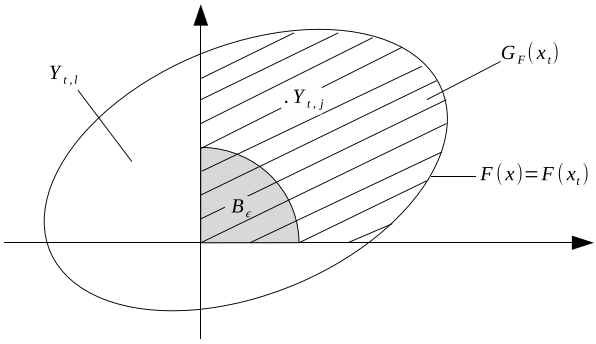


Fig. 7.1 Feasible points, ϵ-optimal points

Remark 7.2 (*Multi-Start Methods*) Instead of the iteration process (7.31a), (7.31a) and corresponding to multi-start methods, cf. [2], we may use the 1-step iterations

$$X_0^* = x_0^* := x_{0,k} \to X_1^* = X_{1,k}^*, \quad k = 1, 2, \ldots, K, \tag{7.31c}$$

based on a random variation of the initial point $x_0 = x_{0,k},\ k = 1, 2, \ldots, K$, according to a certain probability distribution on the feasible domain D.

7.5.2 Probability of Reaching B_ϵ from the Outside

Having a point $X_t^* = x_t^* \notin B_\epsilon$ outside the set B_ϵ of ϵ-optimal points, we consider the probability that the next iteration point X_{t+1}^* is in B_ϵ.

Assuming $F^* + \epsilon < F(x^*)$ and therefore $B_\epsilon \subset G_F(x_t^*)$, we have

$$Y_t^* \in B_\epsilon \iff Y_{t,j} \in B_\epsilon \subset G_F(x_t^*) \text{ for at least one index } j,\ 1 \le j \le r. \tag{7.32a}$$

Thus,

$$P(Y_t^* \notin B_\epsilon \mid x_t^*) = P(Y_{t,1} \notin B_\epsilon, \ldots, Y_{t,r} \notin B_\epsilon \mid x_t^*)$$

and therefore

$$P(Y_t^* \in B_\epsilon \mid x_t^*) = 1 - P(Y_{t,1} \notin B_\epsilon, \ldots, Y_{t,r} \notin B_\epsilon \mid x_t^*). \tag{7.32b}$$

According to the definition (7.31a), (7.31b) of X_{t+1}^*, we then get

$$\begin{aligned} P(X_{t+1}^* \in B_\epsilon \mid x_t^*) &= P(Y_t^* \in B_\epsilon \mid x_t^*) \\ &= 1 - P(Y_{t,j} \notin B_\epsilon,\ 1 \le j \le r \mid x_t^*), \quad x_t^* \notin B_\epsilon. \end{aligned} \tag{7.32c}$$

Hence, from (7.32c) and inequality (7.13) we get the following result.

Lemma 7.4 *Let $x_t^* \notin B\epsilon$ be a realization of an iteration point X_t^* outside the set of ϵ-optimal points B_ϵ.*

(a) If $Y_{t,j},\ 1 \le j \le r$, are stochastically independent search points, given $X_t^* = x_t^*$, then

$$P(X_{t+1}^* \in B_\epsilon \mid x_t^*) = 1 - \prod_{j=1}^{r} P(Y_{t,j} \notin B_\epsilon \mid x_t^*). \tag{7.32d}$$

(b) In case of i.i.d. (independent, identically distributed) variables $Y_{t,j} = Y_t,\ j = 1, \ldots, r$*, it is*

$$P(X^*_{t+1} \in B_\epsilon \mid x^*_t) = 1 - \left(P(Y_t \notin B_\epsilon \mid x^*_t)\right)^r . \tag{7.32e}$$

(c) For i.i.d. variables $Y_{t,j} = Y_t,\ j = 1, \ldots, r$ *we also have*

$$P(X^*_{t+1} \in B_\epsilon) = 1 - \left(1 - P(Y_t \in B_\epsilon \mid x^*_t)\right)^r \geq 1 - e^{r\,P(Y_t \in B_\epsilon \mid x^*_t)} \tag{7.32f}$$

and therefore

$$P(X^*_{t+1} \in B_\epsilon \mid x^*_t) \to 1, \quad r \to \infty. \tag{7.32g}$$

References

1. Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience, Hoboken (2006)
2. Martí, R., et al.: Multi-start methods. In: Martí, R., et al. (eds.) Handbook of Heuristics, pp. 155–175. Springer International Publishing, Cham (2018)

Chapter 8
Approximation of Feedback Control Systems

Abstract Optimal feedback controls under stochastic uncertainty can be obtained in general by approximate methods only. In this chapter approximate feedback controls are obtained by Taylor expansion of the state function with respect to the gain matrix or the gain parameters. Since the state function is the solution of the state equation, hence, a first-order system of differential equations, corresponding systems of differential equations for the partial derivatives of the state function with respect to the gain matrix, the gain parameters, resp., can be obtained by partial differentiation of the state equation with respect to the gain matrix, the gain parameters. Corresponding approximate optimal stochastic feedback control problems are then derived.

8.1 Introduction

In addition to the approximation method based on open-loop feedback control, presented in Sect. 3.2, an approximation procedure for feedback control systems is proposed by using Taylor expansion methods.

Here, we consider nonlinear and linear control systems:

$$\dot{z}(t) = f(t, a, z(t), u(t)), \quad t \geq t_0 \tag{8.1a}$$

$$z(t_0) = z_0(a), \tag{8.1b}$$

and

$$\dot{z}(t) = A(t, a)z(t) + B(t, a)u(t) + c(t, a), \quad t \geq t_0 \tag{8.2a}$$

$$z(t_0) = z_0(a). \tag{8.2b}$$

In (8.1a)–(8.2b), t denotes the continuous time, t_0 the initial time. $z = z(t)$ and $u = u(t)$ are the state n-vector and the control m-vector, resp.. a denotes a possibly random parameter r-vector on a probability space $(\Omega, \mathfrak{A}, \mathcal{P})$. Moreover, $f = f(t, a, z, u)$ is a sufficiently smooth function of (t, a, z, u).

For the linear case $A = A(t, a)$, $B = B(t, a)$ are $n \times n$, $m \times m$ matrices and $c = c(t, a)$ is an n-vector. They all may depend on time t and parameter vector a.

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9_8

Furthermore, we assume that the systems of differential equations (8.1a)–(8.1b) and (8.2a)–(8.2b) have a unique solution

$$z = z(t, t_0, z_0, a, u(\cdot)), \quad t_0 \le t \le t_f, \tag{8.3}$$

on the time intervals $[t_0; t_f]$ under consideration and that this solution is sufficiently differentiable with respect to the parameters arising, e.g., in the feedback control function.

Optimizing the underlying control system (8.1a)–(8.1b) or (8.2a)–(8.2b), one has an objective function $f = F(u(\cdot))$ defined by the total costs

$$F(u(\cdot)) := \int_{t_0}^{t_f} L(t, a, z(t), u(t))\, dt + L_f(t_f, a, z(t_f)) \tag{8.4a}$$

along the trajectory $z = z(t)$ and the terminal point $z_f = z(t_f)$. In case of optimal control systems under stochastic uncertainty, the objective function is defined by the expected total costs

$$F(u(\cdot)) := E\left(\int_{t_0}^{t_f} L(t, a(\omega), z(t, \omega), u(t, \omega))\, dt + L_f(t_f, a(\omega), z(t_f, \omega)) \right). \tag{8.4b}$$

8.2 Control Laws

There are three main types of control functions, see Sect. 3.2:

(i) **open-loop control** $u = u(t),\ t \ge t_0$
Here, the control is a m-vector valued function on the time interval $[t_0; t_f]$ with a certain initial and terminal time t_0, t_f.

(ii) **closed-loop or (state-) feedback control** $u = u(t, z(t)),\ t \ge t_0$
The (state-) feedback control is a function depending on time t and the state $z = z(t)$ arising at time t. More general, the feedback control can be defined by a function of time t and the information $\mathfrak{A}_t$ available up to time t. In many cases linear feedback functions

$$u(t) = u_0(t) + G(t)z(t), \quad t \ge t_0 \tag{8.5}$$

are taken, where $G = G(t)$ is the so-called $n \times m$ *gain matrix* which may depend on time t. Furthermore, $u_0 = u_0(t)$ is a certain function on time t.

(iii) **open-loop feedback control**
This is an approximate feedback control obtained by means of repeated open-loop control:
Based on the time interval $[t_0; t_f]$, consider an *intermediate* time point, selected continuously

$$t_b = t, t_0 \le t < t_f, \tag{8.6a}$$

or selected *stepwise* with a time step $\Delta t > 0$,

$$t_b = t_k := t_0 + k\Delta t < t_f, k = 0, 1, 2, \ldots, \tag{8.6b}$$

and the information $\mathfrak{A}_{t_b}$, known up to time t_b, as, e.g., the state vector $z_b = z(t_b)$ at time t_b. Then, determine the open-loop control

$$u = u_{[t_b;t_f]}(s; t_b; \mathfrak{A}_{t_b}), t_b \le s \le t_f, \tag{8.6c}$$

for the remaining time interval $[t_b; t_f]$. Taking then only the control value at time $s = t_b$, the open-loop feedback control is defined by

$$u_{\mathrm{OLF}}(t_b) := u_{[t_b;t_f]}(t_b; t_f; \mathfrak{A}_{t_b}) \tag{8.6d}$$

continuously, stepwise, resp., for $t_b = t,\ t_0 \le t < t_f$, or $t_b = t_k := t_0 + k\Delta t < t_f,\ k = 0, 1, 2, \ldots$.

8.3 Linear State-Feedback Control Systems

We start our feedback control approximation method with the linear control system (8.2a)–(8.2b) and the state-feedback control (8.5).

Hence, for the state n-vector $z = z(t)$ we have the system of first-order differential equations

$$\dot{z}(t) = A(t)z(t) + B(t)\,(u_0(t) + Gz(t)) + c(t), \quad t \ge t_0. \tag{8.7}$$

The $m \times n$ gain matrix G is represented by

$$G = \sum_{i=1}^{m}\sum_{j=1}^{n} g_{ij} E_{ij} \tag{8.8a}$$

with the elements g_{ij} of the matrix G and the $m \times n$ matrices

$$E_{ij} = \left(e_{lk}^{(ij)}\right)_{l,k=1,\ldots m,n} \tag{8.8b}$$

with

$$e_{lk}^{(ij)} = 1,\ (l,k) = (i,j), \qquad e_{lk}^{(ij)} = 0,\ (l,k) \neq (i,j). \tag{8.8c}$$

Inserting (8.8a) into (8.7), we get

$$\dot{z} = A(t)z + B(t)\left(u_0(t) + \sum_{i=1}^{m}\sum_{j=1}^{n} g_{ij}E_{ij}z\right) + c(t), \quad t \geq t_0. \tag{8.9a}$$

Now it holds

$$E_{ij}z = z_j e_j, \quad i = 1,\dots m, \quad j = 1,\dots,n \tag{8.9b}$$

and with the i-th column e_i of the $m \times m$ unit matrix I and the i-th column b_i of B,

$$B(t)E_{ij}z = B(t)z_j e_i = z_j b_i, \quad i = 1,\dots,m, \quad j = 1,\dots,n. \tag{8.9c}$$

Because of (8.9b)–(8.9c), system (8.9a) can be represented by

$$\dot{z} = A(t)z + B(t)u_0(t) + \sum_{i=1}^{m}\sum_{j=1}^{n} g_{ij}z_j b_i(t) + c(t), \quad t \geq t_0. \tag{8.10}$$

8.3.1 *Taylor Expansion of the Feedback Control System with Respect to the Gain Matrix $G = (g_{ij})$*

Based on calculus, see, e.g., [1, 2], we assume that for the initial state z_0, the prior control function $u_0 = u_0(t)$ and the elements g_{ij} of the gain matrix G under consideration there exists a unique solution

$$z = z(t, t_0, z_0, u_0(\cdot), g_{ij},\ i = 1,\dots m,\ j = 1,\dots n), \quad t \geq t_0 \tag{8.11}$$

of (8.10) on the time interval $[t_0; t_f]$.

Moreover, we assume that the state function (8.11) is sufficiently often differentiable with respect to the gain elements $g_{ij},\ 1 = 1,\dots,m\ j = 1,\dots,n$ and that the time derivative $\frac{d}{dt}$ and the partial derivatives $\frac{\partial}{\partial g_{ij}}$ can be interchanged.

Starting now the expansion of the solution (8.11) of the state equation (8.10) with respect to the elements g_{ij} of G at $G = 0$, for $G = 0$ we get the open-loop system:

$$\dot{z} = A(t)z(t) + B(t)u_0(t) + c(t), \quad t \geq t_0 \tag{8.12a}$$

$$z(t_0) = z_0. \tag{8.12b}$$

If $\Phi = \Phi(t, s)$ denotes the fundamental matrix of the system matrix $A = A(t)$, (8.12a)–(8.12b) has the solution

$$\begin{aligned} z(t) &= z(t, z_0, u_0(\cdot), 0) \\ &= \Phi(t, t_0)z_0 + \int_{t_0}^{t} \Phi(t, s)\left(B(s)u_0(s) + c(s)\right)\, ds, \quad t \geq t_0. \end{aligned} \tag{8.12c}$$

According to the above assumptions, for the partial derivatives

$$\frac{\partial z}{\partial g_{lk}}(t) = \frac{\partial z}{\partial g_{lk}}(t, t_0, z_0, u_0(\cdot), G), \quad l = 1, \ldots, m \;\; j = 1, \ldots, n$$

we get the following systems of differential equations (*variational or perturbation equations*):

$$\frac{d}{dt}\frac{\partial z}{\partial g_{lk}}(t) = A(t)\frac{\partial z}{\partial g_{lk}} + z_k b_l + \sum_{i=1}^{m}\sum_{j=1}^{n} g_{ij}\frac{\partial z_j}{\partial g_{lk}} b_i, \quad t \geq t_0 \tag{8.13a}$$

$$\frac{\partial z}{\partial g_{lk}}(t_0) = 0, \quad l = 1, \ldots m, \;\; k = 1, \ldots n\,. \tag{8.13b}$$

Taking $G = 0$ in (8.13a)–(8.13b), for the partial derivatives

$$\frac{\partial z}{\partial g_{lk}}(t) = \frac{\partial z}{\partial g_{lk}}(t, t_0, z_0, u_0(\cdot), G), \quad l = 1, \ldots m, \;\; j = 1, \ldots n,$$

we have the system of linear differential equations

$$\frac{d}{dt}\frac{\partial z}{\partial g_{lk}}(t) = A(t)\frac{\partial z}{\partial g_{lk}}(t) + z_k(t, t_0, z_0, u_0(t), 0)b_l(t), \quad t \geq t_0 \tag{8.14a}$$

$$\frac{\partial z}{\partial g_{lk}}(t_0) = 0, \tag{8.14b}$$

where function $z_k(t) = z_k(t, t_0, z_0, u_0(t), 0)$ is determined by (8.12c). Corresponding to (8.12c), the solution of (8.14a)–(8.14b) reads

$$\frac{\partial z}{\partial g_{lk}}(t, t_0.z_0, u_0(\cdot), 0) = \int_{t_0}^{t} \Phi(t, s)z_k(s, t_0, z_0, u_0(\cdot), 0)b_l(s)\, ds, \quad t \geq t_0. \tag{8.15}$$

By further partial differentiation of (8.13a)–(8.13b), for the second-order partial derivatives we get

$$\frac{d}{dt}\frac{\partial^2 z}{\partial g_{lk}^2}(t) = A(t)\frac{\partial^2 z}{\partial g_{lk}^2}(t) + \frac{\partial z_k}{\partial g_{lk}} b_l(t) + \sum_{i=1}^{m}\sum_{j=1}^{n} g_{ij}\frac{\partial^2 z_j}{g_{lk}^2} b_i \frac{\partial z_k}{\partial g_{lk}} b_l(t), \qquad t \geq t_0 \tag{8.16a}$$

$$\frac{\partial^2 z}{\partial g_{lk}^2}(t_0) = 0, \tag{8.16b}$$

and for $(u, v) \neq (l, k)$

$$\frac{d}{dt}\frac{\partial z^2}{\partial g_{uv}\partial g_{lk}}(t) = A(t)\frac{\partial^2 z}{\partial g_{uv}\partial g_{lk}}(t) + \frac{\partial z_k}{\partial g_{uv}} b_l(t) + \sum_{i=1}^{m}\sum_{j=1}^{n} g_{ij}\frac{\partial^2 z_j}{\partial g_{uv}\partial g_{lk}}(t) b_i + \frac{\partial z_v}{\partial g_{lk}} b_u(t), \qquad t \geq t_0 \tag{8.16c}$$

$$\frac{\partial^2 z}{\partial g_{uv}\partial g_{lk}}(t_0) = 0. \tag{8.16d}$$

Taking $G = 0$ in (8.16a)–(8.16d) for the derivatives

$$\frac{\partial^2 z}{\partial g_{lk}^2}(t, t_0, z_0, u_0(\cdot), 0), \quad \frac{\partial^2 z}{\partial g_{uv}\partial g_{lk}}(t, t_0, z_0, u_0(\cdot), 0)$$

we have the following systems of linear differential equations:

$$\frac{d}{dt}\frac{\partial^2 z}{\partial g_{lk}^2}(t) = A(t)\frac{\partial z^2}{\partial g_{lk}^2} + 2\frac{\partial z_k}{\partial g_{lk}}(t, t_0, z_0, u_0(\cdot), 0) b_l(t), \quad t \geq t_0 \tag{8.17a}$$

$$\frac{\partial^2 z}{\partial g_{lk}^2}(t_0) = 0 \tag{8.17b}$$

$$\frac{d}{dt}\frac{\partial z^2}{\partial g_{uv}\partial g_{lk}}(t) = A(t)\frac{\partial^2 z}{\partial g_{uv}\partial g_{lk}}(t) + \frac{\partial z_k}{\partial g_{uv}}(t, t_0, z_0, u_0(\cdot), 0) b_l(t) + \frac{\partial z_v}{\partial g_{lk}}(t, t_0, z_0, u_0(\cdot), 0) b_u(t), \qquad t \geq t_0 \tag{8.17c}$$

$$\frac{\partial^2 z}{\partial g_{uv}\partial g_{lk}}(t_0) = 0. \tag{8.17d}$$

Consequently, corresponding to the first-order derivatives of $z = z(t, t_0, z_0, u_0(\cdot), G)$ at $G = 0$ and according to (8.17a)–(8.17d), the second-order derivatives can be represented by

$$\frac{\partial^2 z}{\partial g_{uv}\partial g_{lk}}(t, t_0, z_0, u_0(\cdot), 0)$$
$$= \int_{t_0}^{t} \Phi(t, s)\left(\frac{\partial z_k}{\partial g_{uv}}(s, t_0, z_0, u_0(\cdot), 0)b_l(s)\right.$$
$$\left. + \frac{\partial z_v}{\partial g_{lk}}(s, t_0, z_0, u_0(\cdot), 0)b_u(s)\right) ds, \quad t \geq t_0. \tag{8.18}$$

According to (8.12c), (8.15), (8.18) with further partial differentiation of (8.17a)–(8.17d) and stepwise setting $G = 0$, we get the following result:

Theorem 8.1 *For each order $p = 1, 2, \ldots$ the partial derivatives with respect to the gain elements g_{ij} at $G = 0$ of the state function $z = z(t, t_0, z_0, u_0(\cdot), G)$ of the feedback control system (8.7), (8.10) resp., have an integral representation involving the $p - 1$th order partial derivatives of $z = z(t, t_0, z_0, u_0(\cdot), G)$ with respect to g_{ij} at $G = 0$.*

8.3.2 Time-Dependent Gain Matrices

Considering now time-dependent gain matrices $G = G(t)$, we suppose that $G(t)$ can be represented by

$$G(t) = \sum_{s=0}^{\bar{s}} \tau_s(t) G_s = G_0 + \sum_{s=1}^{\bar{s}} \tau_s(t) G_s \tag{8.19a}$$

with time functions $\tau_s = \tau_s(t),\ s = 0, 1, \ldots, \bar{s}$, and $\tau_0(t) := 1$ as, e.g., for the powers $\tau_s(t) = t^s,\ s = 0, 1, \ldots, \bar{s}$. Moreover, $G_s,\ s, = 0, \ldots, \bar{s}$, are fixed $m \times n$ matrices, represented, cf. (8.8a)–(8.8c), by their elements g_{sij}, hence

$$G_s = \sum_{i=1}^{m}\sum_{j=1}^{n} g_{sij} E_{ij}. \tag{8.19b}$$

Thus, we have

$$G(t) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\sum_{s=0}^{\bar{s}} \tau_s(t) g_{sij}\right) E_{ij}. \tag{8.19c}$$

Inserting (8.19c) into the feedback control system (8.7), due to (8.9b)–(8.9c) and corresponding to case (8.10) with fixed matrices G_s, we get

$$\dot{z} = A(t)z(t) + B(t)u_0(t) + \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\sum_{s=0}^{\overline{s}}\tau_s(t)g_{sij}\right)z_j(t)b_i(t) + c(t), \quad t \geq t_0, \tag{8.20a}$$

$$z(t_0) = z_0. \tag{8.20b}$$

Corresponding to (8.10) and (8.11), we also assume, that for the initial state z_0, the prior control $u_0 = u_0(t)$ and the time-dependent gain matrix $G = G(t)$, (8.20a)–(8.20b) has a unique solution

$$z = z(t, t_0, z_0, u_o(\cdot), g_{sij}, 0 \leq s \leq \overline{s}, 1 \leq i \leq m, 1 \leq j \leq n) \tag{8.21}$$

on the time interval $[t_0; t_f]$ under consideration. Moreover, we suppose, see [1, 2], that the state function (8.21) is sufficiently often differentiable with respect to the gain parameters g_{sij}.

According to the above remarks, also in the time-dependent case, the partial derivatives with respect to the gain parameters g_{sij}, $s = 0, \ldots, \overline{s}, i = 1, \ldots, m, j = 1, \ldots, n$ of the state function (8.21) can be obtained by the same method as for constant gain matrices.

We show this for the first-order derivatives of (8.21) with respect to a gain parameter g_{rlk} with $0 \leq r \leq \overline{s}, 1 \leq l \leq m, 1 \leq k \leq n$.

By partial differentiation of (8.20a)–(8.20b) with respect to g_{rlk} we get

$$\begin{aligned}\frac{d}{dt}\frac{\partial z}{\partial g_{rlk}}(t) &= A(t)\frac{\partial z}{\partial g_{rlk}} + \frac{\partial}{\partial g_{rlk}}\sum_{s=0}^{\overline{s}}\sum_{i=1}^{m}\sum_{j=1}^{n} g_{sij}\left(\tau_s(t)z_j(t)b_i(t)\right)\\ &= A(t)\frac{\partial z}{\partial g_{rlk}} + \tau_r(t)z_k(t)b_l(t)\\ &\quad + \sum_{s=0}^{\overline{s}}\sum_{i=1}^{m}\sum_{j=1}^{n} g_{sij}\frac{\partial}{\partial g_{rlk}}\left(\tau_s(t)z_j(t)b_i(t)\right),\end{aligned} \tag{8.22a}$$

$$\frac{\partial z}{\partial g_{rlk}}(t_0) = 0. \tag{8.22b}$$

Setting now $g_{sij} = 0$ for all indices s, i, j for the partial derivative

$$\frac{\partial z}{\partial g_{rlk}}(t) = \frac{\partial z}{\partial g_{rlk}}(t, t_0.z_0.u_0(\cdot), g_{sij} = 0 \text{ for all } s, i, j), \tag{8.22c}$$

we have the ordinary linear system of differential equations

$$\frac{d}{dt}\frac{\partial z}{\partial g_{rlk}}(t) = A(t)\frac{\partial z}{\partial g_{rlk}} + \tau_r(t) z_k(t) b_l(t), \tag{8.23a}$$

$$\frac{\partial z}{\partial g_{rlk}}(t_0) = 0, \tag{8.23b}$$

with $z_k(t) = z_k(t, t_0, u_0(\cdot), g_{s_{ij}} = 0$ for all $s, i, j)$.

Remark 8.1 In the time-dependent case (8.19a)–(8.19b), the partial derivatives of the state function $z(\cdot)$ at $G = 0$ have the same properties as stated in theorem 8.1 for the time-independent case treated in Sect. 8.3.1.

8.4 Optimal Feedback Control Problem

Based on the Taylor expansion of the state function (8.11), we now present an approximate optimal feedback control problem for the case of a time-independent gain matrix $G = (g_{ij})$, including a first-order approximation of the state function $z(\cdot)$ with respect to G at $G = 0$. Hence, let

$$\begin{aligned} z^{(1)} &= z^{(1)}(t, t_0, z_0, u_o(\cdot), G) \\ &:= z(t, t_0, z_0, u_0(\cdot), 0) + \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{\partial z}{\partial g_{ij}}(t, t_0, z_0, u_0(\cdot), 0) g_{ij} \end{aligned} \tag{8.24a}$$

denote the first-order approximation of the state function (8.11), where the zero- and first-order derivatives at $(t, t_0, z_0, u_0(\cdot), 0)$ are determined by the systems of linear differential equations (8.12a)–(8.12b) and (8.14a)–(8.14b). Moreover, using here the approximate state function (8.24a), the feedback control function (8.5) can be approximated by

$$u^{(1)} = u^{(1)}(t, t_0, z_0, u_0(\cdot), G) := u_0(t) + G z^{(1)}(t, t_0, z_0, u_0(\cdot), G). \tag{8.24b}$$

In the optimal feedback control problem with an objective function (8.4b) we now apply the state and feedback control functions (8.24a)–(8.24b). This yields the following approximate optimal feedback control problem under stochastic uncertainty

$$\min \quad E\left(\int_{t_0}^{t_f} L\left(t, a(\omega), z^{(1)}(t, t_0, z_0(\omega), u_0(\cdot), G), u^{(1)}(t, t_0, z_0(\omega), u_0(\cdot), G)\right) dt\right.$$
$$\left. + L_f\left(t_f, a(\omega), z^{(1)}(t_f, t_0, z_0(\omega), a(\omega), u_0(\cdot), G)\right)\right) \tag{8.25a}$$

s.t.

$$\text{conditions } (8.12a) - (8.12b), (8.14a) - (8.14b) \text{ for } z^{(1)}, u^{(1)} \tag{8.25b}$$
$$u_0(\cdot) \in U, \ G \in \Gamma, \tag{8.25c}$$

where U, Γ resp., denote feasible domains for $u_0(\cdot)$, G.

8.4.1 Stepwise Optimization of $u_0(\cdot)$, G

Solving (8.25a)–(8.25c) approximately, we may first optimize the open-loop control $u_0(\cdot)$ with the corresponding zero-state function

$$z^{(0)} = z^{(0)}(t, t_0, z_0, a, u_0(\cdot)) := z(t, t_0, z_0, s, u_0(\cdot), 0) \tag{8.26}$$

by

Step 1: Optimization of the open-loop control $u_0(\cdot)$ only.
Solve (8.25a)–(8.25c) with the following changes:

(i) In (8.25a), replace $z^{(1)} \to z^{(0)}$, $u^{(1)} \to u_0(\cdot)$,
(ii) in (8.25b), only use (8.12a)–(8.12b),
(iii) in (8.25c), only use "$u_0(\cdot) \in U$".

Let then $u_0^*(\cdot)$ denote the optimal solution of the resulting simplified control problem (8.24a)–(8.25c)$_{u_0(\cdot)}$.

Step 2: Optimization of the gain matrix G Solve (8.25a)–(8.25c) with these changes:

(i) In (8.25a), replace $u_0(\cdot)$ by $u_0^*(\cdot)$,
(ii) in (8.25b), replace $u_0(\cdot)$ by $u_0^*(\cdot)$,
(iii) in (8.25c), only use "$G \in \Gamma$".

Denote then $G^* = (g_{ij}^*)$ the optimal solution of the resulting simplified control problem (8.25a)–(8.25c)$_G$.

Having an optimal solution $(u_0^*(\cdot), G^*)$ of (8.25a)–(8.25c), or an approximate stepwise solution of (8.25a)–(8.25c) as described above, the optimal control of the original problem of minimizing (8.4b) subject to (8.1a)–(8.1b), (8.2a)–(8.2b), resp. and possible constraints for the control $u_0(\cdot)$ and the gain matrix G can be approximated, cf. (8.5), by

$$u^*(t) \approx u_0^*(t) + G^* z(t), \tag{8.27a}$$

where $z = z(t)$ denotes the state vector observed/measured at time t, $t_0 \le t \le t_f$.

Improvements of this method can be obtained by updating $(u_0^*(\cdot), G^*)$ at certain intermediate initial data

$$(t_b, z_b),\ z_b = z(t_b),\ t_0 < t_b < t_f, \tag{8.27b}$$

where z_b denotes the actual state vector at an intermediate time t_b.

8.5 Approximation of Nonlinear Feedback Control Systems

We now consider nonlinear control systems represented, see (8.1a)–(8.1b), by

$$\dot{z}(t) = f(t, a, z(z), u(t)), \qquad t_0 \le t \le t_f, \tag{8.28a}$$

$$z(t_0) = z_0, \tag{8.28b}$$

where the control function $u = u(t)$ is given, cf. (8.5), by

$$u(t) = u_0(t) + Gz(t), \qquad t_0 \le t \le t_f, \tag{8.28c}$$

with an open-loop control $u_0 = u_0(t)$.

Since time-dependent gain matrices $G = G(t)$ can be treated similar to stationary ones, we only consider here time-independent gain matrices represented, see (8.8a), by

$$G = \sum_{i=1}^{m} \sum_{j=1}^{n} g_{ij} E_{ij}. \tag{8.28d}$$

Thus, corresponding to (8.10), in the present case, for the state function

$$z = (t, t_0, z_0, a, u_0(\cdot), g_{ij}, i = 1, \ldots, m,\ j = 1, \ldots, n), \quad t_0 \le t \le t_f, \tag{8.28e}$$

we have, see (8.9b), the following system of differential equations

$$\begin{aligned} \dot{z}(t) &= f(t, a, z(t), u_0(t) + Gz(t)) \\ &= f\left(t, a, z(t), u_0(t) + \sum_{i=1}^{m} \sum_{j=1}^{n} g_{ij} z_j(t) e_i\right), \quad t_0 \le t \le t_f, \end{aligned} \tag{8.29a}$$

$$z(t_0) = z_0, \tag{8.29b}$$

where e_i denotes again the i-th column of the $m \times m$ unit matrix.

Corresponding to (8.13a)–(8.13b), by differentiation of (8.29a)–(8.29b) with respect to a gain parameter g_{lk}, for the partial derivative of the state function (8.28e) with respect to g_{lk} we get the system of linear differential equations

$$\frac{d}{dt}\frac{\partial z}{\partial g_{lk}}(t) = \frac{\partial f}{\partial z}(t, a, z(t), u(t))\frac{\partial z}{\partial g_{lk}}(t) + \frac{\partial f}{\partial u}(t, a, z(t), u(t))\left(z_k(t)e_l + G\frac{\partial z}{\partial g_{lk}}(t)\right) \tag{8.30a}$$

$$\frac{\partial z}{\partial g_{lk}}(t_0) = 0, \qquad l = 1, \ldots, m, \ k = 1, \ldots, n. \tag{8.30b}$$

Taking $G = 0$, for the derivative

$$\frac{\partial z}{\partial g_{lk}} = \frac{\partial z}{\partial g_{lk}}(t, t_0, z_0, a, u_0(\cdot), 0)$$

we get the system of linear differential equations

$$\frac{d}{dt}\frac{\partial z}{\partial glk}(t) = \frac{\partial f}{\partial z}(t, a, z(t), u_0(t))\frac{\partial z}{\partial g_{lk}}(t) + \frac{\partial f}{\partial u}(t, a, z(t), u_0(t))z_k(t)e_l, \quad t_0 \le t \le t_f \tag{8.31a}$$

$$\frac{\partial z}{\partial g_{lk}}(t_0) = 0, \tag{8.31b}$$

where $z(t) = z_1(t), \ldots, z_n(t))^T = z(t, t_0, z_0, a, u_0(\cdot), 0)$ is determined by the systems of differential equations

$$\dot{z}(t) = f(t, a, z(t), u_0(t)), \quad t_0 \le t \le t_f, \tag{8.31c}$$

$$z(t_0) = z_0. \tag{8.31d}$$

Remark 8.2 Second and higher order derivatives of $z = z(t)$, see (8.28a)–(8.28e), with respect to g_{ij}, $i = 1, \ldots, m$ $j = 1, \ldots, n$ can be obtained as shown in case of state linear differential equations.

8.6 Approximation Error

According to the properties of parameter-dependent systems of differential equations, see [1, 2], under weak assumptions, the unique solution

$$z = \left(z_k(t, t_0, z_0, u_0(\cdot), g_{ij}, \ i = 1, \ldots, m, \ j = 1, \ldots, n)\right)_{1 \le k \le n} \tag{8.32a}$$

of (8.1a)–(8.1b), (8.2a)–(8.2b), (8.28a)–(8.28b), (8.29a)–(8.29b), resp., is a sufficiently differentiable n-vector function of, among other variables, the gain parameters

$$g = (g_1, g_2, \ldots, g_{m\cdot n})^T := (g_{11}, \ldots, g_{1n}, \ldots, g_{m1}, \ldots, g_{mn})^T. \tag{8.32b}$$

Hence, the accuracy of a linear or higher order approximation of the state variables $z_k = z_k(t),\ k = 1, \ldots, n$ by the Taylor polynomials with respect to g,

$$\begin{aligned} T_p(t) &= T_p(t, t_0, z_0, u_0(\cdot), g) \\ &:= z(t, t_0, z_0, u_0(\cdot), 0) + \sum_{l=1}^{m\cdot n} \frac{\partial z}{\partial g_l}(t, t_0, z_0, u_0(\cdot), 0) g_l + \cdots \\ &+ \frac{1}{p!} \sum_{l_1, l_2, \ldots, l_n = 1}^{m\cdot n} \frac{\partial^p z}{\partial g_{l_1} \partial g_{l_2} \ldots \partial g_{l_n}}(t, t_0, z_0, u_0(\cdot), 0) \prod_{l=1}^{m\cdot n} g_l \end{aligned} \tag{8.32c}$$

can be evaluated by means of the Taylor formula. The corresponding Lagrange remainder term reads

$$R_p(t, g, 0) := \frac{1}{(p+1)!} \sum_{l_1, l_2, \ldots, l_{n+1}}^{m\cdot n} \frac{\partial^{p+1} z}{\partial g_{l_1} \partial g_{l_2} \ldots \partial g_{l_{n+1}}}(t, t_0, z_0, u_0(\cdot), \vartheta_g) \prod_{i=1}^{p+1} g_{l_i} \tag{8.32d}$$

with $0 < \vartheta < 1$.

Remark 8.3

(a) The sums and products in (8.32c)–(8.32d) can be represented component-wise for the components $z_k,\ k = 1, 2, \ldots, n$ of the state vector function $z = z(t)$.
(b) The derivatives $\frac{\partial^p z}{\partial g_{l_1} \ldots \partial g_{l_p}}$ are build, see Theorem 8.1, in nested way by the corresponding lower order, $p-1, p-2, \ldots, 0$, derivatives up to the state function $z = z(t, t_0, z_0, u_0(\cdot), g)$.

8.7 Extensions

In this Section, first special representations of the open-loop or prior control function $u_0 = u_0(t)$ are presented. Moreover, generalizations of the approximation method for feedback control systems with linear to nonlinear feedback functions of the state $z = z(t)$ are presented.

8.7.1 Special Representations of the Open-Loop (Prior) Control Function $u_0(\cdot)$

Solving optimal (feedback) control problems as considered in Sect. 8.6, advantages can be obtained if the open-loop control m-vector function $u_0 = u_0(t)$ is represented by

$$u_0(t) = \sum_{l=1}^{L} \mathcal{T}_l(t) u_l, \tag{8.33a}$$

where

$$\mathcal{T}_l(t) = \begin{pmatrix} \tau_{l11}(t) & \tau_{l12}(t) & \dots & \tau_{l1m}(t) \\ \vdots & \vdots & & \vdots \\ \tau_{lm1}(t) & \tau_{lm2}(t) & \dots & \tau_{lmm}(t) \end{pmatrix}, \quad t \geq t_0,\ l = 1, \dots, L, \tag{8.33b}$$

are (m, m)-matrix functions involving given time functions $\tau_{lij} = \tau_{lij}(t)$, $i, j = 1, \dots, m$, $l = 1, \dots, L$, as, e.g., certain powers of time t.

Furthermore, u_l, $l = 1, \dots, L$, are unknown m-vectors to be determined optimally.

8.7.1.1 Linear State Equations

In case of optimal open-loop control problems with linear state equations, cf. (8.12a), (8.12b) and Sect. 8.4, the state function $z(t) = z(t, t_0, z_0, u_0(\cdot))$ is given by

$$z(t, t_0, z_0, u_0(\cdot)) = \Phi(t, t_0) z_0 + \int_{t_0}^{t} \Phi(t, s) \left(B(s) u_0(s) + c(s) \right) ds,$$

where, with the fundamental matrix $Y = Y(t)$ of the system matrix $A = A(t)$, it holds $\Phi(t, s) = Y(t) Y(s)^{-1}$.

Using now definition (8.33a), (8.33b) of $u_0(\cdot)$, we find

$$z(t, t_0, z_0, u_0(\cdot)) = \Phi(t, t_0) z_0 + \sum_{l=1}^{L} \left(\int_{t_0}^{t} \Phi(t, s) B(s) T_l(s)\, ds \right) u_l + \int_{t_0}^{t} \Phi(t, s) c(s)\, ds. \tag{8.34a}$$

Thus, the state function $z = z(t)$ can be represented as an affine-linear function

$$z(t) = z(t, t_0, z_0, u_1, \dots, u_L) \tag{8.34b}$$

of the parameter m-vectors u_l, $l = 1, \dots, L$.

This yields considerable simplifications, e.g., in endpoint control problems and trajectory optimization problems, see Chapters 9 and 10.

8.7.2 Nonlinear Feedback Function

Generalizing the control function (8.5) with a linear feedback term $u_F(t,z) = G(t)z$, we consider here control functions with nonlinear feedback laws

$$u(t,z) = u_0(t) + G(t,z). \tag{8.35a}$$

Here, $G = G(t,z)$ is defined by

$$G(t,z) := \sum_{l=1}^{L} g_l G_l(t,z) \tag{8.35b}$$

with given m-vector functions $G_l = G_l(t,z)$, $l = 1,\ldots,L$, being nonlinear in z, and scalar parameters g_l, $l = 1,\ldots,L$. A further representation is

$$G(t,z) := \sum_{l=1}^{L} G_l(t,z) g_l \tag{8.35c}$$

with matrix functions $G_l = G_l(t,z)$, being nonlinear in state z, and corresponding vectorial parameters g_l, $l = 1,\ldots,L$. Using the feedback representation defined by (8.35a)–(8.35c), the state function $z(t) = z(t,t_0,z_0,u(\cdot,\cdot))$ is reduced to

$$z = z(t,t_0,z_0,u_0(\cdot),g_1,\ldots,g_L) \tag{8.36a}$$

involving the open-loop control $u_0(\cdot)$ and the parameter vector

$$g := (g_1,g_2,\ldots,g_L)^T, \quad g = (g_1^T,g_2^T,\ldots,g_L^T)^T \text{ resp..} \tag{8.36b}$$

For simplicity, consider only linear systems of differential equations for the state function (8.36a), (8.36b). In case of scalar feedback parameters $g_1,\ldots,g_L$, we have the differential equation

$$\frac{dz}{dt}(t) = At)z(t) + B(t)\left(u_0(t) + \sum_{l=1}^{L} g_l G_l(t,z(t))\right) + c(t) \tag{8.37a}$$

$$z(t_0) = z_0. \tag{8.37b}$$

Using the method suggested above for linear feedback, under corresponding assumptions, Taylor expansions of the state function (8.36a), (8.36b) with respect to g at $g = 0$ can be derived.

For $g = 0$ we start again with the solution

$$z(t) = z(t, t_0, z_0, u_0(\cdot), 0) \tag{8.38}$$

of the open-loop state equation (8.12a), (8.12b). Corresponding to (8.13a), (8.13b), for the partial derivatives of the state function (8.36a) with respect to $g_l, l = 1, \ldots, L$, by partial differentiation of system (8.37a), (8.37b) with respect to $g_l,\ l = 1, \ldots, L$, we get the systems

$$\frac{d}{dt}\frac{\partial z}{\partial g_l}(t) = A(t)\frac{\partial z}{\partial g_l}(t) + B(t)\left(G_l(t, z(t)) + \sum_{\lambda=1}^{L} g_\lambda \frac{\partial G_\lambda}{\partial z}(t, z(t))\frac{\partial z}{\partial g_l}(t),\right) \tag{8.39a}$$

$$\frac{\partial z}{\partial g_l}(t_0) = 0. \tag{8.39b}$$

Taking now $g = 0$, for the partial derivatives

$$\frac{\partial z}{\partial g_l} = \frac{\partial z}{\partial g_l}(t, t_0, z_0, u_0(\cdot), 0), \quad l = 1, \ldots L,$$

we get the linear differential equations

$$\frac{d}{dt}\frac{\partial z}{\partial g_l}(t) = A(t)\frac{\partial z}{\partial g_l}(t) + B(t)G_l(t, z(t, t_0, z_0, u(\cdot), 0)), \quad t \geq 0 \tag{8.40a}$$

$$\frac{\partial z}{\partial g_l}(t_0) = 0. \tag{8.40b}$$

Remark 8.4

(a) Obviously, similar systems of differential equations for higher order partial derivatives of (8.36a) with respect to the parameters $g_l,\ l = 1, \ldots, L$ at $g = 0$ can be obtained.

(b) Furthermore, comparing the above system of differential equations for the case of nonlinear feedback with the related systems (8.7), (8.13a), (8.13b), (8.14a), (8.14b) for linear feedback, we find that they have the same structures concerning the sequence of higher partial derivatives, cf. Theorem 8.1.

References

1. Dieudonné, J.: Foundations of Modern Analysis. Academic, New York (1969)
2. Walter, W.: Gewöhnliche Differentialgleichungen. Springer, Berlin (2000)

Chapter 9
Stochastic Optimal Open-Loop Feedback Control

Abstract In this chapter a second procedure for an approximate determination of stochastic optimal feedback controls is based on the stochastic open-loop feedback method. This very efficient approximation method is also the basis of the model predictive control procedures. Using the methods mentioned in Chap. 3, stochastic optimal open-loop feedback controls are constructed by computing next to stochastic optimal open-loop controls on the *remaining time intervals* $t_b \le t \le t_f$ with $t_0 \le t_b \le t_f$. Having stochastic optimal open-loop feedback controls on each remaining time interval $t_b \le t \le t_f$ with $t_0 \le t_b \le t_f$, a stochastic optimal open-loop feedback control law follows then immediately by evaluating each of the stochastic optimal open-loop controls on $t_b \le t \le t_f$ at the corresponding initial time point $t = t_b$. The efficiency of this method has been proved already by applications to the stochastic optimization of regulators for robots.

9.1 Dynamic Structural Systems Under Stochastic Uncertainty

9.1.1 Stochastic Optimal Structural Control: Active Control

In order to omit structural damages and therefore high compensation (recourse) costs, active control techniques are used in structural engineering. The structures usually are stationary, safe, and stable without considerable external dynamic disturbances. Thus, in case of heavy dynamic external loads, such as earthquakes, wind turbulences, water waves, etc., which cause large vibrations with possible damages, additional control elements can be installed in order to counteract applied dynamic loads, see [3, 18, 19].

The structural dynamics is modeled mathematically by means of a linear system of second-order differential equations for the $m-$vector $q = q(t)$ of displacements. The system of differential equations involves random dynamic parameters, random initial values, the random dynamic load vector, and a control force vector depending on an input control function $u = u(t)$. Robust, i.e., parameter-insensitive optimal feedback controls u^* are determined in order to cope with the stochastic uncertainty

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9_9

involved in the dynamic parameters, the initial values, and the applied loadings. In practice, the design of controls is directed often to reduce the mean square response (displacements and their time derivatives) of the system to a desired level within a reasonable span of time.

The performance of the resulting structural control problem under stochastic uncertainty is evaluated therefore by means of a convex quadratic cost function $L = L(t, z, u)$ of the state vector $z = z(t)$ and the control input vector $u = u(t)$. While the actual time path of the random external load is not known at the planning stage, we may assume that the probability distribution or at least the moments under consideration of the applied load and other random parameters are known. The problem is then to determine a robust, i.e., parameter-insensitive (open-loop) feedback control law by minimization of the expected total costs, hence, a stochastic optimal control law.

As mentioned above, in active control of dynamic structures, cf. [3, 14, 18–22], the behavior of the m-vector $q = q(t)$ of displacements with respect to time t is described by a system of second-order linear differential equations for $q(t)$ having a right-hand side being the sum of the stochastic applied load process and the control force depending on a control n-vector function $u(t)$:

$$M\ddot{q} + D\dot{q} + Kq(t) = f\Big(t, \omega, u(t)\Big), \quad t_0 \le t \le t_f\,. \tag{9.1a}$$

Hence, the force vector $f = f\Big(t, \omega, u(t)\Big)$ on the right-hand side of the dynamic equation (9.1a) is given by the sum

$$f(t, \omega, u) = f_0(t, \omega) + f_a(t, \omega, u) \tag{9.1b}$$

of the applied load $f_0 = f_0(t, \omega)$ being a vector valued stochastic process describing, e.g., external loads or excitation of the structure caused by earthquakes, wind turbulences, water waves, etc., and the actuator or control force vector $f_a = f_a(t, \omega, u)$ depending on an input or control n-vector function $u = u(t)$, $t_0 \le t \le t_f$. Here, ω denotes the random element, lying in a certain probability space (Ω, A, P), used to represent random variations. Furthermore, M, D, K, resp., denotes the $m \times m$ mass, damping and stiffness matrix. In many cases the actuator or control force f_a is linear, i.e.,

$$f_a = \Gamma_u u \tag{9.1c}$$

with a certain $m \times n$ matrix Γ_u.

By introducing appropriate matrices, the linear system of second-order differential equations (9.1a), (9.1b) can be represented by a system of first-order differential equations as follows:

$$\dot{z} = g(t, \omega, z(t, \omega), u) := Az(t, \omega) + Bu + b(t, \omega) \tag{9.2a}$$

with

$$A := \begin{pmatrix} 0 & I \\ -M^{-1}K & -M^{-1}D \end{pmatrix}, \qquad B := \begin{pmatrix} 0 \\ M^{-1}\Gamma_u \end{pmatrix}, \tag{9.2b}$$

$$b(t,\omega) := \begin{pmatrix} 0 \\ M^{-1} f_0(t,\omega). \end{pmatrix} \tag{9.2c}$$

Moreover, $z = z(t)$ is the $2m$-state vector defined by

$$z = \begin{pmatrix} q \\ \dot{q} \end{pmatrix} \tag{9.2d}$$

fulfilling a certain initial condition

$$z(t_0) = \begin{pmatrix} q(t_0) \\ \dot{q}(t_0) \end{pmatrix} := \begin{pmatrix} q_0 \\ \dot{q}_0 \end{pmatrix} \tag{9.2e}$$

with given or stochastic initial values $q_0 = q_0(\omega), \dot{q}_0 = \dot{q}_0(\omega)$.

9.1.2 *Stochastic Optimal Design of Regulators*

In the optimal design of regulators for dynamic systems, see also Chap. 10, the (m-) vector $q = q(t)$ of *tracking errors* is described by a system of $2nd$ order linear differential equations:

$$M(t)\ddot{q} + D(t)\dot{q} + K(t)q(t) = -Y(t)\Delta p_D(\omega) + \Delta u(t,\omega), \quad t_0 \le t \le t_f. \tag{9.3}$$

Here, $M(t), D(t), K(t), Y(t)$ denote certain time-dependent Jacobians arising from the linearization of the dynamic equation around the stochastic optimal reference trajectory and the conditional expectation $\overline{p_D}$ of the vector of dynamic parameters $p_D(\omega)$. The deviation between the vector of dynamic parameters $p_D(\omega)$ and its conditional expectation $\overline{p_D}$ is denoted by $\Delta p_D(\omega) := p_D(\omega) - \overline{p_D}$. Furthermore, $\Delta u(t)$ denotes the correction of the feedforward control $u^0 = u^0(t)$.

By introducing appropriate matrices, system (9.3) can be represented by the $1st$ order system of linear differential equations:

$$\dot{z} = A(t)z(t,\omega) + B\Delta u + b(t,\omega) \tag{9.4a}$$

with

$$A(t) := \begin{pmatrix} 0 & I \\ -M(t)^{-1}K & -M(t)^{-1}D(t) \end{pmatrix}, \qquad B := \begin{pmatrix} 0 \\ M(t)^{-1} \end{pmatrix}, \tag{9.4b}$$

$$b(t,\omega) := \begin{pmatrix} 0 \\ -M(t)^{-1}Y(t)\Delta p_D(\omega) \end{pmatrix}. \tag{9.4c}$$

Again, the $(2m$-) state vector $z = z(t)$ is defined by

$$z = \begin{pmatrix} q \\ \dot{q} \end{pmatrix}. \tag{9.4d}$$

9.1.3 Robust (Optimal) Open-Loop Feedback Control

According to the description in Sect. 3.2, a feedback control is defined, cf. (3.10b), by

$$u(t) := \varphi(t, \mathcal{I}_t), t \geq t_0, \tag{9.5a}$$

where $\mathcal{I}_t$ denotes again the total information about the control system up to time t and $\varphi(\cdot,\cdot)$ designates the feedback control law. If the state $z_t := z(t)$ is available at each time point t, the control input n-vector function $u = u(t)$, $\Delta u = \Delta u(t)$, resp., can be generated by means of a PD -controller, hence,

$$u(t)(\Delta u(t)) := \varphi\Big(t, z(t)\Big), \quad t \geq t_0, \tag{9.5b}$$

with a feedback control law $\varphi = \varphi(t, q, \dot{q}) = \varphi(t, z(t))$. Efficient approximate feedback control laws are constructed here by using the concept of **open-loop feedback control**. Open-loop feedback control is the main tool in *model predictive control*, cf. [1, 8, 16], which is very often used to solve optimal control problems in practice. The idea of *open-loop feedback control* is to construct a feedback control law quasi *argument-wise*, see cf. [2, 5].

A major issue in optimal control is the **robustness**, cf. [4], i.e., the insensitivity of the optimal control with respect to parameter variations. In case of random parameter variations, robust optimal controls can be obtained by means of stochastic optimization methods, cf. [10]. Thus, we introduce the following concept of an *stochastic optimal (open-loop) feedback control*.

Definition 9.1 In case of stochastic parameter variations, robust, hence, parameter-insensitive optimal (open-loop) feedback controls obtained by stochastic optimization methods are also called **stochastic optimal (open-loop) feedback controls**.

9.1.4 Stochastic Optimal Open-Loop Feedback Control

Finding a stochastic optimal open-loop feedback control, hence, an optimal (open-loop) feedback control law, see Sect. 3.2, being insensitive as far as possible with respect to random parameter variations, means that besides optimality of the control law also its insensitivity with respect to stochastic parameter variations should be guaranteed. Hence, in the following sections we develop now a stochastic version of the (optimal) open-loop feedback control method, cf. [9, 11–13]. A short overview on this novel stochastic optimal open-loop feedback control concept is given below.

At each intermediate time point $t_b \in [t_0, t_f]$, based on the information $\mathcal{I}_{t_b}$ available at time t_b, a stochastic optimal open-loop control $u^* = u^*\big(t; t_b, \mathcal{I}_{t_b}\big),\ t_b \le t \le t_f$, is determined first on the remaining time interval $[t_b, t_f]$, see Fig. 9.1, by stochastic optimization methods, cf. [10].

Having a stochastic optimal open-loop control $u^* = u^*\big(t; t_b, \mathcal{I}_{t_b}\big),\ t_b \le t \le t_f$, on each remaining time interval $[t_b, t_f]$ with an arbitrary starting time $t_b,\ t_0 \le t_b \le t_f$, a stochastic optimal open-loop feedback control law is then defined, see Definition 3.2, as follows:

Definition 9.2

$$\varphi^* = \varphi\big(t_b, \mathcal{I}_{t_b}\big) := u^*(t_b) = u^*\big(t_b; t_b, \mathcal{I}_{t_b}\big), \quad t_0 \le t_b \le t_f. \tag{9.5c}$$

Hence, at time $t = t_b$ just the "first" control value $u^*(t_b) = u^*\big(t_b; t_b, \mathcal{I}_{t_b}\big)$ of $u^*\big(\cdot; t_b, \mathcal{I}_{t_b}\big)$ is used only. For each other argument $\big(t, \mathcal{I}_t\big)$ the same construction is applied.

For finding stochastic optimal open-loop controls, based on the methods developed in Chap. 3, on the remaining time intervals $t_b \le t \le t_f$ with $t_0 \le t_b \le t_f$, the stochastic Hamilton function of the control problem is introduced. Then, the class of $H-$ minimal controls, cf. Definitions 3.6 and 3.7, can be determined in case of stochastic uncertainty by solving a finite-dimensional stochastic optimization problem for minimizing the conditional expectation of the stochastic Hamiltonian subject to the remaining deterministic control constraints at each time point t. Having a $H-$ minimal control, the related two-point boundary value problem with random parameters can be formulated for the computation of a stochastic optimal state- and costate trajectory. Due to the linear-quadratic structure of the underlying control problem, the state and costate trajectory can be **determined analytically** to a large extent. Inserting then these trajectories into the H-minimal control, stochastic optimal open-loop controls are found on an arbitrary remaining time interval. According to

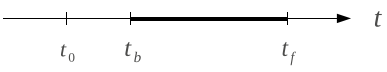


Fig. 9.1 Remaining time interval

Definition 9.2, these controls yield then immediately a stochastic optimal open-loop feedback control law. Moreover, the obtained controls can be realized in **real-time**, which is already shown for applications in optimal control of industrial robots, cf. [17].

Summarizing, we get *optimal (open-loop) feedback controls under stochastic uncertainty* minimizing the effects of external influences on system behavior, subject to the constraints of not having a complete representation of the system, cf. [4]. Hence, robust or stochastic optimal active controls are obtained by **new** techniques from *Stochastic Optimization*, see [10]. Of course, the construction can be applied also to $PD-$ and $PID-$controllers.

9.2 Expected Total Cost Function

The performance function F for active structural control systems is defined, cf. [6–8], by the conditional expectation of the total costs being the sum of costs L along the trajectory, arising from the displacements $z = z(t, \omega)$ and the control input $u = u(t, \omega)$, and possible terminal costs G arising at the final state z_f. Hence, on the remaining time interval $t_b \le t \le t_f$ we have the following conditional expectation of the total cost function with respect to the information $\mathfrak{A}_{t_b}$ available up to time t_b:

$$F := E\left(\int_{t_b}^{t_f} L\Big(t, \omega, z(t, \omega), u(t, \omega)\Big)dt + G(t_f, \omega, z(t_f, \omega))\Big|\, \mathfrak{A}_{t_b}\right). \tag{9.6a}$$

Supposing quadratic costs along the trajectory, the function L is given by

$$L(t, \omega, z, u) := \frac{1}{2} z^T Q(t, \omega) z + \frac{1}{2} u^T R(t, \omega) u \tag{9.6b}$$

with positive (semi) definite $2m \times 2m$, $n \times n$, resp., matrix functions $Q = Q(t, \omega)$, $R = R(t, \omega)$. In the simplest case the weight matrices Q, R are fixed. A special selection for Q reads

$$Q = \begin{pmatrix} Q_q & 0 \\ 0 & Q_{\dot{q}} \end{pmatrix} \tag{9.6c}$$

with positive (semi) definite weight matrices Q_q, $Q_{\dot{q}}$, resp., for $q, \dot{q}$. Furthermore, $G = G(t_f, \omega, z(t_f, \omega))$ describes possible terminal costs. In case of endpoint control G is defined by

$$G(t_f, \omega, z(t_f, \omega)) := \frac{1}{2}(z(t_f, \omega) - z_f(\omega))^T G_f (z(t_f, \omega) - z_f(\omega)), \tag{9.6d}$$

where $G_f = G_f(\omega)$ is a positive (semi) definite, possible random weight matrix, and $z_f = z_f(\omega)$ denotes the (possible random) final state.

Remark 9.1 Instead of $\frac{1}{2}u^T Ru$, in the following we also use a more general convex control cost function $C = C(u)$.

9.3 Open-Loop Control Problem on the Remaining Time Interval $[t_b, t_f]$

In the following we suppose next to that the $2m \times 2m$ matrix A and the $2m \times n$ matrix B are given, fixed matrices.

Having the differential equation with random coefficients derived above, describing the behavior of the dynamic mechanical structure/system under stochastic uncertainty, and the costs arising from displacements and at the terminal state, on a given remaining time interval $[t_b, t_f]$ a stochastic optimal open-loop control $u^* = u^*\big(t; t_b, \mathcal{I}_{t_b}\big)$, $t_b \le t \le t_f$, is a solution of the following optimal control problem under stochastic uncertainty:

$$\min \quad E\left(\int_{t_b}^{t_f} \frac{1}{2}\left(z(t,\omega)^T Qz(t,\omega) + u(t)^T Ru(t)\right)\, dt + G(t_f, \omega, z(t_f,\omega))\Big|\mathfrak{A}_{t_b}\right) \tag{9.7a}$$

$$\text{s.t.} \quad \dot{z}(t,\omega) = Az(t,\omega) + Bu(t) + b(t,\omega), \text{ a.s.}, \qquad t_b \le t \le t_f \tag{9.7b}$$

$$z(t_b,\omega) = \overline{z_b}^{(b)} \quad \text{(estimated state at time } t_b\text{)} \tag{9.7c}$$

$$u(t) \in D_t, \qquad t_b \le t \le t_f\,. \tag{9.7d}$$

An important property of (9.7a)–(9.7d) is stated next:

Lemma 9.1 *If the terminal cost function $G = G(t_f, \omega, z)$ is convex in z, and the feasible domain D_t is convex for each time point t, $t_0 \le t \le t_f$, then the stochastic optimal control problem (9.7a)–(9.7d) is a convex optimization problem.*

9.4 The Stochastic Hamiltonian of (9.7a)–(9.7d)

According to (3.28a), (3.40a), see also [8], the stochastic Hamiltonian H related to the stochastic optimal control problem (9.7a)–(9.7d) reads

$$\begin{aligned} H(t,\omega,z,y,u) &:= L(t,\omega,z,u) + y^T g(t,\omega,z,u) \\ &= \frac{1}{2}z^T Qz + C(u) + y^T\left(Az + Bu + b(t,\omega)\right)\,. \end{aligned} \tag{9.8a}$$

9.4.1 Expected Hamiltonian (with Respect to the Time Interval $[t_b, t_f]$ and Information $\mathfrak{A}_{t_b}$)

For the definition of a $H-$minimal control the conditional expectation of the stochastic Hamiltonian is needed:

$$\begin{aligned}\overline{H}^{(b)} := E\left(H(t,\omega,z,y,u)\big|\mathfrak{A}_{t_b}\right) &= E\left(\frac{1}{2}z^T Qz + y^T\left(Az + b(t,\omega)\right)\big|\mathfrak{A}_{t_b}\right)\\ &\quad + C(u) + E\left(y^T Bu\big|\mathfrak{A}_{t_b}\right)\\ &= C(u) + E\left(B^T y(t,\omega)\big|\mathfrak{A}_{t_b}\right)^T u + \cdots = C(u) + h(t)^T u + \ldots\end{aligned} \tag{9.8b}$$

with

$$h(t) = h\left(t; t_b, \mathcal{I}_{t_b}\right) := E\left(B(\omega)^T y(t,\omega)\big|\mathfrak{A}_{t_b}\right), \quad t \geq t_b\,. \tag{9.8c}$$

9.4.2 H-Minimal Control on $[t_b, t_f]$

In order to formulate the two-point boundary value problem for a stochastic optimal open-loop control $u^* = u^*\left(t; t_b, \mathcal{I}_{t_b}\right)$, $t_b \leq t \leq t_f$, we need first an H-minimal control

$$\widetilde{u^*} = \widetilde{u^*}\left(t, z(t,\cdot), y(t,\cdot); t_b, \mathcal{I}_{t_b}\right), \quad t_b \leq t \leq t_f,$$

defined, see Definitions 3.6 and 3.7 and cf. also [8], for $t_b \leq t \leq t_f$ as a solution of the following convex stochastic optimization problem, cf. [10]:

$$\min E\left(H(t,\omega,z(t,\omega),y(t,\omega),u)\big|\mathfrak{A}_{t_b}\right) \tag{9.9a}$$

s.t.

$$u \in D_t, \tag{9.9b}$$

where $z = z(t,\omega)$, $y = y(t,\omega)$ are certain trajectories.

According to (9.9a), (9.9b) the H-minimal control

$$\widetilde{u}^* = \widetilde{u}^*\Big(t, z(t,\cdot), y(t,\cdot); t_b, \mathcal{I}_{t_b}\Big) = \widetilde{u}^*(t, h(\cdot; t_b, \mathcal{I}_{t_b})) \tag{9.10a}$$

is defined by

$$\widetilde{u}^*(t, h(\cdot; t_b, \mathcal{I}_{t_b})) := \operatorname*{argmin}_{u \in D_t} C(u) + h\big(t; t_b, \mathcal{I}_{t_b}\big)^T u \qquad \text{for } t \geq t_b\,. \tag{9.10b}$$

9.4.2.1 Strictly Convex Cost Function, no Control Constraints

For strictly convex, differentiable cost functions $C = C(u)$, as, e.g., $C(u) = \frac{1}{2}u^T Ru$ with positive definite matrix R, the necessary and sufficient condition for $\widetilde{u}^*$ reads in case $D_t = \mathbb{R}^n$:

$$\nabla C(u) + h\big(t; t_b, \mathcal{I}_{t_b}\big) = 0\,. \tag{9.11a}$$

If $u \mapsto \nabla C(u)$ is a 1-1-operator, then the solution of (9.11a) reads

$$u = v(h\big(t; t_b, \mathcal{I}_{t_b}\big)) := \nabla C^{-1}(-h\big(t; t_b, \mathcal{I}_{t_b}\big))\,. \tag{9.11b}$$

With (9.8c) and (9.10b) we then have

$$\widetilde{u}^*(t,h) = \widetilde{u}^*(h\big(t; t_b, \mathcal{I}_{t_b}\big))) := \nabla C^{-1}\left(-E\left(B(\omega)^T y(t,\omega)\big|\mathfrak{A}_{t_b}\right)\right)\,. \tag{9.11c}$$

9.5 Canonical (Hamiltonian) System

We suppose here that a H-minimal control $\widetilde{u}^* = \widetilde{u}^*\big(t, z(t,\cdot), y(t,\cdot); t_b, \mathcal{I}_{t_b}\big)$, $t_b \leq t \leq t_f$, i.e., a solution $\widetilde{u}^* = \widetilde{u}^*(t,h) = v(h(t)))$ of the stochastic optimization problem (9.9a), (9.9b) is available. Moreover, the conditional expectation $E\left(\xi\big|\mathfrak{A}_{t_b}\right)$ of a random variable ξ is also denoted by $\overline{\xi}^{(b)}$, cf. (9.8b). According to (3.46), Theorem 3.7, a stochastic optimal open-loop control $u^* = u^*(t; t_b, \mathcal{I}_{t_b})$, $t_b \leq t \leq t_f$,

$$u^*(t; t_b, \mathcal{I}_{t_b}) = \widetilde{u}^*\Big(t, z^*(t,\cdot), y^*(t,\cdot); t_b, \mathcal{I}_{t_b}\Big), \quad t_b \leq t \leq t_f, \tag{9.12}$$

of the stochastic optimal control problem (9.7a)–(9.7d), can be obtained, see also [8], by solving the following stochastic two-point boundary value problem related to (9.7a)–(9.7d).

Theorem 9.1 *If $z^* = z^*(t,\omega)$, $y^* = y^*(t,\omega)$, $t_0 \le t \le t_f$, is a solution of*

$$\dot{z}(t,\omega) = Az(t,\omega) + B\nabla C^{-1}\left(-\overline{B^T y(t)}^{(b)}\right) + b(t,\omega), \quad t_b \le t \le t_f \tag{9.13a}$$

$$z(t_b,\omega) = \overline{z_b}^{(b)} \tag{9.13b}$$

$$\dot{y}(t,\omega) = -A^T y(t,\omega) - Qz(t,\omega) \tag{9.13c}$$

$$y(t_f,\omega) = \nabla G(t_f,\omega,z(t_f,\omega)), \tag{9.13d}$$

then the function $u^* = u^*(t;t_b,\mathcal{I}_{t_b})$, $t_b \le t \le t_f$, defined by (9.12) is a stochastic optimal open-loop control for the remaining time interval $t_b \le t \le t_f$.

9.6 Minimum-Energy Control

In this case we have $Q = 0$, i.e., there are no costs for the displacements $z = \begin{pmatrix} q \\ \dot{q} \end{pmatrix}$. In this case the solution of (9.13c), (9.13d) reads

$$y(t,\omega) = e^{A^T(t_f-t)}\nabla_z G(t_f,\omega,z(t_f,\omega)), \qquad t_b \le t \le t_f\,. \tag{9.14a}$$

This yields for fixed Matrix B

$$\widetilde{u^*}(t,h(t)) = v(h(t)) = \nabla C^{-1}\left(-B^T e^{A^T(t_f-t)}\overline{\nabla_z G(t_f,z(t_f))}^{(b)}\right),$$
$$t_b \le t \le t_f. \tag{9.14b}$$

Having (9.14a), (9.14b), for the state trajectory $z = z(t,\omega)$ we get, see (9.13a), (9.13b), the following system of ordinary differential equations

$$\dot{z}(t,\omega) = Az(t,\omega) + B\nabla C^{-1}\big(-B^T e^{A^T(t_f-t)}\overline{\nabla_z G(t_f,z(t_f))}^{(b)}\big) + b(t,\omega), \qquad t_b \le t \le t_f, \tag{9.15a}$$

$$z(t_b,\omega) = \overline{z_b}^{(b)}. \tag{9.15b}$$

The solution of system (9.15a), (9.15b) reads

$$z(t,\omega) = e^{A(t-t_b)}\overline{z_b}^{(b)} + \int_{t_b}^{t} e^{A(t-s)}\Big(b(s,\omega) + B\nabla C^{-1}\left(-B^T e^{A^T(t_f-s)}\overline{\nabla_z G(t_f,z(t_f))}^{(b)}\right)\Big)\,ds,$$
$$t_b \le t \le t_f\,. \tag{9.16}$$

For the final state $z = z(t_f, \omega)$ we get the relation:

$$z(t_f,\omega) = e^{A(t_f-t_b)}\overline{z_b}^{(b)} + \int_{t_b}^{t_f} e^{A(t_f-s)}\Big(b(s,\omega) + B\nabla C^{-1}\left(-B^T e^{A^T(t_f-s)}\overline{\nabla_z G(t_f, z(t_f))}^{(b)}\right)\Big)\,ds\,. \tag{9.17}$$

9.6.1 Endpoint Control

In the case of endpoint control, the terminal cost function is given by the following definition (9.18a), where $z_f = z_f(\omega)$ denotes the desired—possible random—final state:

$$G(t_f,\omega,z(t_f,\omega)) := \frac{1}{2}\|z(t_f,\omega) - z_f(\omega)\|^2\,. \tag{9.18a}$$

Hence,

$$\nabla G(t_f,\omega,z(t_f,\omega)) = z(t_f,\omega) - z_f(\omega) \tag{9.18b}$$

and therefore

$$\begin{aligned}\overline{\nabla G(t_f, z(t_f))}^{(b)} &= \overline{z(t_f)}^{(b)} - \overline{z_f}^{(b)} \\ &= E\left(z(t_f,\omega)\middle|\mathfrak{A}_{t_b}\right) - E\left(z_f(\omega)\middle|\mathfrak{A}_{t_b}\right)\,.\end{aligned} \tag{9.18c}$$

Thus

$$z(t_f,\omega) = e^{A(t_f-t_b)}\overline{z_b}^{(b)} + \int_{t_b}^{t_f} e^{A(t_f-s)}\Big(b(s,\omega) + B\nabla C^{-1}\left(-B^T e^{A^T(t_f-s)}\left(\overline{z(t_f)}^{(b)} - \overline{z_f}^{(b)}\right)\right)\Big)\,ds\,. \tag{9.19a}$$

Taking expectations $E(\ldots|\mathfrak{A}_{t_b})$ in (9.19a), we get the following condition for $\overline{z(t_f)}^{(b)}$:

$$\begin{aligned}\overline{z(t_f)}^{(b)} =& e^{A(t_f-t_b)}\overline{z_b}^{(b)} + \int_{t_b}^{t_f} e^{A(t_f-s)}\overline{b(s)}^{(b)}\,ds \\ &+ \int_{t_b}^{t_f} e^{A(t_f-s)}B\nabla C^{-1}\left(-B^T e^{A^T(t_f-s)}\left(\overline{z(t_f)}^{(b)} - \overline{z_f}^{(b)}\right)\right)\,ds\,.\end{aligned} \tag{9.19b}$$

9.6.1.1 Quadratic Control Costs

Here, the control cost function $C = C(u)$ reads

$$C(u) = \frac{1}{2} u^T R u \,, \tag{9.20a}$$

hence,

$$\nabla C = Ru \tag{9.20b}$$

and therefore

$$\nabla C^{-1}(w) = R^{-1} w \,. \tag{9.20c}$$

Consequently, (9.19b) reads

$$\begin{aligned} \overline{z(t_f)}^{(b)} &= e^{A(t_f - t_b)} \overline{z_b}^{(b)} + \int_{t_b}^{t_f} e^{A(t_f - s)} \overline{b(s)}^{(b)} \, ds \\ &\quad - \int_{t_b}^{t_f} e^{A(t_f - s)} B R^{-1} B^T e^{A^T (t_f - s)} \, ds \overline{z(t_f)}^{(b)} \\ &\quad + \int_{t_b}^{t_f} e^{A(t_f - s)} B R^{-1} B^T e^{A^T (t_f - s)} \, ds \overline{z_f}^{(b)} \,. \end{aligned} \tag{9.21}$$

Define now

$$U := \int_{t_b}^{t_f} e^{A(t_f - s)} B R^{-1} B^T e^{A^T (t_f - s)} \, ds \,. \tag{9.22}$$

Lemma 9.2 *$I + U$ is regular.*

Proof Due to the previous considerations, U is a positive semidefinite $2m \times 2m$ matrix. Hence, U has only nonnegative eigenvalues.

Assuming that the matrix $I + U$ is singular, there is a $2m$-vector $w \neq 0$ such that

$$(I + U)\, w = 0.$$

However, this yields

$$Uw = -Iw = -w = (-1)w,$$

which means that $\lambda = -1$ is an eigenvalue of U. Since this contradicts to the above mentioned property of U, the matrix $I + U$ must be regular. □

From (9.21) we get

$$(I+U)\,\overline{z(t_f)}^{(b)} = e^{A(t_f-t_b)}\overline{z_b}^{(b)} + \int_{t_b}^{t_b} e^{A(t_f-s)}\overline{b(s)}^{(b)}\,ds + U\overline{z_f}^{(b)}, \tag{9.23a}$$

hence,

$$\begin{aligned}\overline{z(t_f)}^{(b)} &= (I+U)^{-1}\,e^{A(t_f-t_b)}z_b + (I+U)^{-1}\int_{t_b}^{t_b} e^{A(t_f-s)}\overline{b(s)}^{(b)}\,ds \\ &\quad + (I+U)^{-1}\,U\overline{z_f}^{(b)}\,.\end{aligned} \tag{9.23b}$$

Now, (9.23b) and (9.18b) yield

$$\begin{aligned}\overline{\nabla_z G(t_f, z(t_f))} &= \overline{z(t_f) - z_f}^{(b)} = \overline{z(t_f)}^{(b)} - \overline{z_f}^{(b)} \\ &= (I+U)^{-1}\,e^{A(t_f-t_b)}\overline{z_b}^{(b)} \\ &\quad + (I+U)^{-1}\int_{t_b}^{t_f} e^{A(t_f-s)}\overline{b(s)}^{(b)}\,ds \\ &\quad + \big((I+U)^{-1}\,U - I\big)\,\overline{z_f}^{(b)}\,.\end{aligned} \tag{9.24}$$

Thus, a stochastic optimal open-loop control $u^* = u^*(t; t_b, \mathcal{I}_{t_b})$, $t_b \le t \le t_f$, on $[t_b, t_f]$ is given by, cf. (9.11b),

$$\begin{aligned}u^*(t; t_b, \mathcal{I}_{t_b}) = -R^{-1}B^T e^{A^T(t_f-t)}\Big(&(I+U)^{-1}\,e^{A(t_f-t_b)}\overline{z_b}^{(b)} \\ &+ (I+U)^{-1}\int_{t_b}^{t_f} e^{A(t_f-s)}\overline{b(s)}^{(b)}\,ds \\ &+ \big((I+U)^{-1}\,U - I\big)\,\overline{z_f}^{(b)}\Big), \qquad t_b \le t \le t_f.\end{aligned} \tag{9.25}$$

Finally, the stochastic optimal open-loop feedback control law $\varphi = \varphi(t, \mathcal{I}_t)$ is then given by

$$\begin{aligned}\varphi(t_b, \mathcal{I}_{t_b}) &:= u^*(t_b; t_b, \mathcal{I}_{t_b}) \\ &= -R^{-1}B^T e^{A^T(t_f-t_b)}\,(I+U)^{-1}\,e^{A(t_f-t_b)}\overline{z_b}^{(b)} \\ &\quad - R^{-1}B^T e^{A^T(t_f-t_b)}\,(I+U)^{-1}\int_{t_b}^{t_f} e^{A(t_f-s)}\overline{b(s)}^{(b)}\,ds \\ &\quad - R^{-1}B^T e^{A^T(t_f-t_b)}\left((I+U)^{-1}U - I\right)\overline{z_f}^{(b)} \end{aligned} \tag{9.26}$$

with $\mathcal{I}_{t_b} := \left(\overline{z_b}^{(b)} := \overline{z(t_b)}^{(b)}, \overline{b(\cdot)}^{(b)}, \overline{z_f}^{(b)}\right)$.

Replacing $t_b \to t$, we find this result:

Theorem 9.2 *The stochastic optimal open-loop feedback control law $\varphi = \varphi(t, \mathcal{I}_t)$ is given by*

$$\begin{aligned}\varphi(t, \mathcal{I}_t) = &\underbrace{-R^{-1}B^T e^{A^T(t_f-t)}\,(I+U)^{-1}\,e^{A(t_f-t)}}_{\Psi_0(t)}\overline{z(t)}^{(t)} \\ &\underbrace{-R^{-1}B^T e^{A^T(t_f-t)}\,(I+U)^{-1}\int_t^{t_f} e^{A(t_f-s)}\overline{b(s)}^{(t)}\,ds}_{\Psi_1(t,\overline{b(\cdot)}^{(t)})} \\ &\underbrace{-R^{-1}B^T e^{A^T(t_f-t)}\left((I+U)^{-1}U - I\right)}_{\Psi_2(t)}\overline{z_f}^{(t)}\,, \end{aligned} \tag{9.27a}$$

hence,

$$\begin{aligned}\varphi(t, \mathcal{I}_t) &= \Psi_0(t)\overline{z(t)}^{(t)} + \Psi_1(t, \overline{b(\cdot)}^{(t)}) + \Psi_2(t)\overline{z_f}^{(t)}\,, \\ \mathcal{I}_t &:= \left(\overline{z(t)}^{(t)}, \overline{b(\cdot)}^{(t)}, \overline{z_f}^{(t)}\right). \end{aligned} \tag{9.27b}$$

Remark 9.2 Note that the stochastic optimal open-loop feedback law $\overline{z(t)}^{(t)} \mapsto \varphi(t, \mathcal{I}_t)$ is not linear in general, but affine linear.

9.6.2 Endpoint Control with Different Cost Functions

In this section we consider more general terminal cost functions G. Hence, suppose

$$G(t_f, \omega, z(t_f, \omega)) := \kappa(z(t_f, \omega) - z_f(\omega))\,, \tag{9.28a}$$

$$\nabla G(t_f, \omega, z(t_f, \omega)) = \nabla\kappa(z(t_f, \omega) - z_f(\omega))\,. \tag{9.28b}$$

Consequently,

$$\widetilde{u^*}(t, h(t)) = v^*(h(t)) = \nabla C^{-1} \left(B^T e^{A^T (t_f - t)} \overline{\nabla \kappa (z(t_f) - z_f}^{(b)} \right) \tag{9.29a}$$

and therefore, see (9.17)

$$\begin{aligned} z(t_f, \omega) &= e^{A(t_f - t_b)} z_b + \int_{t_b}^{t_f} e^{A(t_f - s)} b(s, \omega)\, ds \\ &\quad + \int_{t_b}^{t_f} e^{A(t_f - s)} B \nabla C^{-1} \left(-B^T e^{A^T (t_f - s)} \overline{\nabla \kappa (z(t_f) - z_f}^{(b)} \right) ds\,, \\ &\quad t_b \le t \le t_f. \end{aligned} \tag{9.29b}$$

Special case:

Now a special terminal cost function is considered in more detail:

$$\kappa(z - z_f) := \sum_{i=1}^{2m} (z_i - z_{f_i})^4 \tag{9.30a}$$

$$\nabla \kappa (z - z_f) = 4 \left((z_1 - z_{f_1})^3, \ldots, (z_{2m} - z_{f_{2m}})^3 \right)^T . \tag{9.30b}$$

Here,

$$\begin{aligned} \overline{\nabla \kappa (z - z_f)}^{(b)} &= 4 \left(E \left((z_1 - z_{f_1})^3 | \mathfrak{A}_{t_b} \right), \ldots, E \left((z_{2m} - z_{f_{2m}})^3 | \mathfrak{A}_{t_b} \right) \right)^T \\ &= 4 \left(m_3^{(b)}(z_1(t_f, \cdot); z_{f_1}(\cdot)), \ldots, m_3^{(b)}(z_{2m}(t_f, \cdot); z_{f_{2m}}(\cdot)) \right)^T \\ &=: 4 m_3^{(b)}(z(t_f, \cdot); z_f(\cdot))\,. \end{aligned} \tag{9.31}$$

Thus,

$$\begin{aligned} z(t_f, \omega) &= e^{A(t_f - t_b)} z_b + \int_{t_b}^{t_f} e^{A(t_f - s)} b(s, \omega)\, ds \\ &\quad + \underbrace{\int_{t_b}^{t_f} e^{A(t_f - s)} B \nabla C^{-1} \left(-B^T e^{A^T (t_f - s)} 4 m_3^{(b)}(z(t_f, \cdot); z_f(\cdot)) \right) ds}_{J\left(m_3^{(b)}(z(t_f, \cdot); z_f(\cdot)) \right)}\,. \end{aligned} \tag{9.32}$$

Equation (9.32) yields then

$$\begin{aligned}&\left(z(t_f,\omega)-z_f(\omega)\right)^3\Big|_{c-by-c}\\&=\left(e^{A(t_f-t_b)}z_b-z_f+\int_{t_b}^{t_f}e^{A(t_f-s)}b(s,\omega)\,ds+J\left(m_3^{(b)}(z(t_f,\cdot);z_f(\cdot))\right)\right)^3_{c-by-c},\end{aligned}\tag{9.33a}$$

where "*c-by-c*" means "*component-by-component*". Taking expectations in (9.33a), we get the following relation for the moment vector $m_3^{(b)}$:

$$m_3^{(b)}(z(t_f,\cdot);z_f(\cdot))=\Psi\left(m_3^{(b)}(z(t_f,\cdot);z_f(\cdot))\right).\tag{9.33b}$$

Remark 9.3

$$\begin{aligned}&E\left(\left(z(t_f,\omega)-z_f(\omega)\right)^3|\mathfrak{A}_{t_b}\right)\Big|_{c-by-c}\\&=E^{(b)}\left(z(t_f,\omega)-\overline{z}^{(b)}(t_f)+\overline{z}^{(b)}(t_f)-z_f(\omega)\right)^3\\&=E^{(b)}\Big(\left(z(t_f,\omega)-\overline{z}^{(b)}(t_f)\right)^3+3\left(z(t_f,\omega)-\overline{z}^{(b)}(t_f)\right)^2\left(\overline{z}^{(b)}(t_f)-z_f(\omega)\right)\\&+3\left(z(t_f,\omega)-\overline{z}^{(b)}(t_f)\right)\left(\overline{z}^{(b)}(t_f)-z_f(\omega)\right)^2+\left(\overline{z}^{(b)}(t_f)-z_f(\omega)\right)^3\Big).\end{aligned}\tag{9.33c}$$

Assuming that $z(t_f,\omega)$ and $z_f(\omega)$ are stochastic independent, then

$$\begin{aligned}&E\left((z(t_f,\omega)-z_f(\omega))^3|\mathfrak{A}_{t_b}\right)\\&=m_3^{(b)}(z(t_f,\cdot))+3\sigma^{2(b)}(z(t_f,\cdot))(\overline{z}^{(b)}(t_f)-\overline{z_f}^{(b)})+\overline{\left(\overline{z}^{(b)}(t_f)-z_f\right)^3}^{(b)},\end{aligned}\tag{9.33d}$$

where $\sigma^{2(b)}(z(t_f,\cdot)$ denotes the conditional variance of the state reached at the final time point t_f, given the information at time t_b.

9.6.3 Weighted Quadratic Terminal Costs

With a certain (possibly random) weight matrix $\Gamma=\Gamma(\omega)$, we consider the following terminal cost function:

$$G(t_f,\omega,z(t_f,\omega)):=frac12\left\|\Gamma(\omega)\left(z(t_f,\omega)-z_f(\omega)\right)\right\|^2.\tag{9.34a}$$

This yields

$$\nabla G(t_f, \omega, z(t_f, \omega)) = \Gamma(\omega)^T \Gamma(\omega)(z(t_f, \omega) - z_f(\omega)) \,, \tag{9.34b}$$

and from (9.14a) we get

$$\begin{aligned} y(t, \omega) &= e^{A^T(t_f - t)} \nabla_z G(t_f, \omega, z(t_f, \omega)) \\ &= e^{A^T(t_f - t)} \Gamma(\omega)^T \Gamma(\omega)(z(t_f, \omega) - z_f(\omega)) \,, \end{aligned} \tag{9.35a}$$

hence,

$$\begin{aligned} \overline{y}^{(b)}(t) &= e^{A^T(t_f - t)} \overline{\Gamma^T \left(\Gamma z(t_f) - \Gamma z_f\right)}^{(b)} \\ &= e^{A^T(t_f - t)} \left(\overline{\Gamma^T \Gamma z(t_f)}^{(b)} - \overline{\Gamma^T \Gamma z_f}^{(b)} \right) . \end{aligned} \tag{9.35b}$$

Thus, for the $H-$ minimal control we find

$$\begin{aligned} \widetilde{u^*}(t, h) &= v(h(t)) \\ &= \nabla C^{-1} \left(-B^T \overline{y}^{(b)}(t) \right) \\ &= \nabla C^{-1} \Bigg(- B^T e^{A^T(t_f - t)} \Big(\overline{\Gamma^T \Gamma z(t_f)}^{(b)} \\ &\qquad - \overline{\Gamma^T \Gamma z_f}^{(b)} \Big) \Bigg) . \end{aligned} \tag{9.36}$$

We obtain therefore, see (9.16),

$$\begin{aligned} z(t, \omega) = e^{A(t - t_b)} z_b + \int_{t_b}^{t} e^{A(t-s)} \Bigg(& b(s, \omega) \\ + B \nabla C^{-1} \Bigg(& - B^T e^{A^T(t_f - s)} \Big(\overline{\Gamma^T \Gamma z(t_f)}^{(b)} - \overline{\Gamma^T \Gamma z_f}^{(b)} \Big) \Bigg) \Bigg) \, ds. \end{aligned} \tag{9.37a}$$

9.6.3.1 Quadratic Control Costs

Assume that the control costs and their gradient are given by

$$C(u) = \frac{1}{2} u^T R u, \ \nabla C(u) = Ru \,. \tag{9.37b}$$

Here, (9.37a) yields

$$z(t_f,\omega) = e^{A(t_f-t_b)}z_b + \int_{t_b}^{t_f} e^{A(t_f-s)}\Big(b(s,\omega) - BR^{-1}B^T e^{A^T(t_f-s)}\left(\overline{\Gamma^T\Gamma z(t_f}^{(b)} - \overline{\Gamma^T\Gamma z_f}^{(b)}\right)\Big)\,ds\,. \tag{9.37c}$$

Multiplying with $\Gamma(\omega)^T\Gamma(\omega)$ and taking expectations, from (9.37c) we get

$$\begin{aligned}\overline{\Gamma^T\Gamma z(t_f)}^{(b)} &= \overline{\Gamma^T\Gamma}^{(b)} e^{A(t_f-t_b)}\overline{z_b}^{(b)} + \int_{t_b}^{t_f} \overline{\Gamma^T\Gamma e^{A(t_f-s)}b(s}^{(b)}\,ds \\ &\quad - \overline{\Gamma^T\Gamma}^{(b)} \int_{t_b}^{t_f} e^{A(t_f-s)}BR^{-1}B^T e^{A^T(t_f-s)}\,ds \\ &\quad \times \left(\overline{\Gamma^T\Gamma z(t_f)}^{(b)} - \overline{\Gamma^T\Gamma z_f}^{(b)}\right). \end{aligned} \tag{9.38a}$$

According to a former lemma, we define the matrix

$$U = \int_{t_b}^{t_f} e^{A(t_f-s)}BR^{-1}B^T e^{A^T(t_f-s)}\,ds\,.$$

From (9.38a) we get then

$$\begin{aligned}&\left(I + \overline{\Gamma^T\Gamma}^{(b)}U\right)\overline{\Gamma^T\Gamma z(t_f)}^{(b)} \\ &= \overline{\Gamma^T\Gamma}^{(b)} e^{A(t_f-t_b)}\overline{z_b}^{(b)} + \int_{t_b}^{t_f} \overline{\Gamma^T\Gamma e^{A(t_f-s)}b(s)}^{(b)}\,ds \\ &+ \overline{\Gamma^T\Gamma}^{(b)}U\overline{\Gamma^T\Gamma z_f}^{(b)}. \end{aligned} \tag{9.38b}$$

Lemma 9.3 *$I + \overline{\Gamma^T\Gamma}^{(b)}U$ is regular.*

Proof First notice that not only U, but also $\overline{\Gamma^T\Gamma}^{(b)}$ is positive semidefinite:

$$v^T\overline{\Gamma^T\Gamma}^{(b)}v = \overline{v^T\Gamma^T\Gamma v} = \overline{(\Gamma v)^T\,\Gamma v}^{(b)} = \overline{\|\Gamma v\|_2^2}^{(b)} \geq 0\,.$$

Then their product $\overline{\Gamma^T\Gamma}^{(b)}U$ is positive semidefinite as well. This follows immediately from [15] as $\Gamma(\omega)^T\Gamma(\omega)$ is symmetric. □

Since the matrix $I + \overline{\Gamma^T\Gamma}^{(b)}U$ is regular, we get cf. (9.23a), (9.23b),

$$\begin{aligned}\overline{\Gamma^T\Gamma z(t_f)}^{(b)} &= \left(I + \overline{\Gamma^T\Gamma}^{(b)}U\right)^{-1}\overline{\Gamma^T\Gamma}^{(b)}e^{A(t_f-t_b)}\overline{z_b}^{(b)}\\ &\quad+\left(I + \overline{\Gamma^T\Gamma}^{(b)}U\right)^{-1}\int_{t_b}^{t_f}\overline{\Gamma^T\Gamma e^{A(t_f-s)}b(s)}^{(b)}\,ds\\ &\quad+\left(I + \overline{\Gamma^T\Gamma}^{(b)}U\right)^{-1}\overline{\Gamma^T\Gamma}^{(b)}U\overline{\Gamma z_f}^{(b)}. \qquad (9.38c)\end{aligned}$$

Putting (9.38c) into (9.36), corresponding to (9.25) we get the stochastic optimal open-loop control

$$\begin{aligned}u^*(t;t_b,\mathcal{I}_{t_b}) &= -R^{-1}B^T e^{A^T(t_f-t)}\left(\overline{\Gamma^T\Gamma z(t_f)}^{(b)} - \overline{\Gamma^T\Gamma z_f}^{(b)}\right)\\ &= \ldots, \quad t_b \le t \le t_f\,, \qquad (9.39)\end{aligned}$$

which yields then the related stochastic optimal open-loop feedback control $\varphi = \varphi(t,\mathcal{I}_t)$ law corresponding to Theorem 9.2.

9.7 Nonzero Costs for Displacements

Suppose here that $Q \neq 0$. According to (9.13a)–(9.13d), for the adjoint trajectory $y = y(t,\omega)$ we have the system of differential equations

$$\begin{aligned}\dot{y}(t,\omega) &= -A^T y(t,\omega) - Qz(t,\omega)\\ y(t_f,\omega) &= \nabla G(t_f,\omega,z(t_f,\omega))\,,\end{aligned}$$

which has the following solution for given $z(t,\omega)$ and $\nabla G(t_f,\omega,z(t_f,\omega))$:

$$y(t,\omega) = \int_t^{t_f} e^{A^T(s-t)}Qz(s,\omega)\,ds + e^{A^T(t_f-t)}\nabla G(t_f,\omega,z(t_f,\omega)). \qquad (9.40)$$

Indeed, we get

$$
\begin{aligned}
y(t_f,\omega) &= 0 + I\nabla_z G(t_f,\omega,z(t_f,\omega)) = \nabla_z G(t_f,\omega,z(t_f,\omega))\\
\dot{y}(t,\omega) &= -e^{A^T\cdot 0}Qz(t,\omega)\\
&\quad - \int_t^{t_f} A^T e^{A^T(s-t)}Qz(s,\omega)\,ds - A^T e^{A^T(t_f-t)}\nabla G(t_f,\omega,z(t_f,\omega))\\
&= -e^{A^T\cdot 0}Qz(t,\omega)\\
&\quad - A^T\left(\int_t^{t_f} e^{A^T(s-t)}Qz(s,\omega)\,ds + e^{A^T(t_f-t)}\nabla G(t_f,\omega,z(t_f,\omega))\right)\\
&= -A^T y(t,\omega) - Qz(t,\omega)\,.
\end{aligned}
$$

From (9.40) we then get

$$
\begin{aligned}
\overline{y}^{(b)}(t) &= E^{(b)}(y(t,\omega)) = E\left(y(t,\omega)|\mathfrak{A}_{t_b}\right)\\
&= \int_t^{t_f} e^{A^T(s-t)}Q\overline{z}^{(b)}(s)\,ds + e^{A^T(t_f-t)}\overline{\nabla G(t_f,z(t_f))}^{(b)}\,.
\end{aligned}
\tag{9.41}
$$

The unknown function $\overline{z(t)}^{(b)}$ and the vector $z(t_f,\omega)$ in this equation are both given, based on $\overline{y(t)}^{(b)}$, by the initial value problem, see (9.13a), (9.13b),

$$
\dot{z}(t,\omega) = Az(t,\omega) + B\nabla C^{-1}\left(-B^T\overline{y}^{(b)}(t)\right) + b(t,\omega) \tag{9.42a}
$$

$$
z(t_b,\omega) = z_b\,. \tag{9.42b}
$$

Taking expectations, considering the state vector at the final time point t_f, resp., yields the expressions:

$$
\overline{z}^{(b)}(t) = e^{A(t-t_b)}\overline{z_b}^{(b)} + \int_{t_b}^{t} e^{A(t-s)}\left(\overline{b(s)}^{(b)} + B\nabla C^{-1}\left(-B^T\overline{y}^{(b)}(s)\right)\right)\,ds, \tag{9.43a}
$$

$$
z(t_f,\omega) = e^{A(t_f-t_b)}z_b + \int_{t_b}^{t_f} e^{A(t_f-s)}\left(b(s,\omega) + B\nabla C^{-1}\left(-B^T\overline{y}^{(b)}(s)\right)\right)\,ds. \tag{9.43b}
$$

9.7.1 Quadratic Control and Terminal Costs

Corresponding to (9.18a), (9.18b) and (9.20a), (9.20b), suppose

$$\nabla G(t_f, \omega, z(t_f, \omega)) = z(t_f, \omega) - z_f(\omega)\,,$$
$$\nabla C^{-1}(w) = R^{-1} w\,.$$

According to (9.12) and (9.11c), in the present case the stochastic optimal open-loop control is given by

$$u^*(t; t_b, \mathcal{I}_{t_b}) = \widetilde{u^*}(t, h(t)) = R^{-1}\left(-E\left(B^T y(t,\omega)\middle|\mathfrak{A}_{t_b}\right)\right) = -R^{-1}B^T\overline{y(t)}^{(b)}. \tag{9.44a}$$

Hence, we need the function $\overline{y}^{(b)} = \overline{y(t)}^{(b)}$. From (9.41) and (9.18a), (9.18b) we have

$$\overline{y(t)}^{(b)} = e^{A^T(t_f-t)}\left(\overline{z(t_f)}^{(b)} - \overline{z_f}^{(b)}\right) + \int_t^{t_f} e^{A^T(s-t)} Q\overline{z(s)}^{(b)}\,ds\,. \tag{9.44b}$$

Inserting (9.43a), (9.43b) into (9.44b), we have

$$\begin{aligned}\overline{y(t)}^{(b)} = e^{A^T(t_f-t)}\Bigg(& e^{A(t_f-t_b)}\overline{z_b}^{(b)} - \overline{z_f}^{(b)} \\ &+ \int_{t_b}^{t_f} e^{A(t_f-s)}\left(\overline{b(s)}^{(b)} - BR^{-1}B^T\overline{y(s)}^{(b)}\right)ds\Bigg) \\ &+ \int_t^{t_f} e^{A^T(s-t)}Q\Bigg(e^{A(s-t_b)}\overline{z_b}^{(b)} \\ &+ \int_{t_b}^{s} e^{A(s-\tau)}\left(\overline{b(\tau)}^{(b)} - BR^{-1}B^T\overline{y(\tau)}^{(b)}\right)d\tau\Bigg)ds\,.\end{aligned} \tag{9.44c}$$

In the following we develop a condition that guarantees the existence and uniqueness of a solution $\overline{y}^b = \overline{y(t)}^{(b)}$ of equation (9.44c).

Theorem 9.3 *In the space of continuous functions, the above Eq. (9.44c) has a unique solution if*

$$c_B < \frac{1}{c_A\sqrt{c_{R^{-1}}(t_f - t_0)\left(1 + \frac{(t_f-t_0)c_Q}{2}\right)}}\,. \tag{9.45}$$

Here,

$$c_A := \sup_{t_b \le t \le s \le t_f} \|e^{A(t-s)}\|_F \quad c_B := \|B\|_F \quad c_{R^{-1}} := \|R^{-1}\|_F \quad c_Q := \|Q\|_F,$$

and the index F denotes the Frobenius-Norm.

Proof The proof of the existence and uniqueness of such a solution is based on the Banach fixed point theorem. For applying this theorem, we consider the Banach space

$$\mathcal{X} = \left\{ f : [t_b; t_f] \to \mathbb{R}^{2m} \ : \ f \text{ continuous} \right\} \tag{9.46a}$$

equipped with the supremum norm

$$\|f\|_L := \sup_{t_b \le t \le t_f} \|f(t)\|_2 \,, \tag{9.46b}$$

where $\|\cdot\|_2$ denotes the Euclidean norm on $\mathbb{R}^{2m}$.

Now we study the operator $\mathcal{T} : \mathcal{X} \to \mathcal{X}$ defined by

$$\begin{aligned}
(\mathcal{T} f)(t) =& e^{A^T(t_f - t)} \Bigg(e^{A(t_f - t_b)} \overline{z_b}^{(b)} - \overline{z_f}^{(b)} \\
&+ \int_{t_b}^{t_f} e^{A(t_f - s)} \left(\overline{b(s)}^{(b)} - B R^{-1} B^T f(s) \right) ds \Bigg) \\
&+ \int_{t}^{t_f} e^{A^T(s-t)} Q \Bigg(e^{A(s-t_b)} \overline{z_b}^{(b)} \\
&+ \int_{t_b}^{s} e^{A(s-\tau)} \left(\overline{b(\tau)}^{(b)} - B R^{-1} B^T f(\tau) \right) d\tau \Bigg) ds \,.
\end{aligned} \tag{9.47}$$

The norm of the difference $\mathcal{T} f - \mathcal{T} g$ of the images of two different elements $f, g \in \mathcal{X}$ with respect to $\mathcal{T}$ may be estimated as follows:

$$\begin{aligned}
&\|\mathcal{T} f - \mathcal{T} g\| \\
&= \sup_{t_b \le t \le t_f} \Bigg\{ \Bigg\| e^{A^T(t_f - t)} \int_{t_b}^{t_f} e^{A(t_f - s)} B R^{-1} B^T (g(s) - f(s))\, ds \\
&+ \int_{t}^{t_f} e^{A^T(s-t)} Q \int_{t_b}^{s} e^{A(s-\tau)} B R^{-1} B^T (g(\tau) - f(\tau))\, d\tau\, ds \Bigg\|_2 \Bigg\} .
\end{aligned} \tag{9.48a}$$

Note that the Frobenius norm is submultiplicative and compatible with the Euclidian norm. Using these properties, we get

$$
\begin{aligned}
&\|\mathcal{T}f - \mathcal{T}g\| \\
&\leq \sup_{t_b \leq t \leq t_f} \left\{ c_A \int_{t_b}^{t_f} c_A c_B c_{R^{-1}} c_B \|f(s) - g(s)\|_2 \, ds \right. \\
&\quad \left. + c_A c_Q \int_{t}^{t_f} \int_{t_b}^{s} c_A c_B c_{R^{-1}} c_B \|f(\tau) - g(\tau)\|_2 \, d\tau \, ds \right\} \\
&\leq \sup_{t_b \leq t \leq t_f} \left\{ c_A \int_{t_b}^{t_f} c_A c_B c_{R^{-1}} c_B \sup_{t_b \leq t \leq t_f} \|f(s) - g(s)\|_2 \, ds \right. \\
&\quad \left. + c_A c_Q \int_{t}^{t_f} \int_{t_b}^{s} c_A c_B c_{R^{-1}} c_B \sup_{t_b \leq t \leq t_f} \|f(\tau) - g(\tau)\|_2 \, d\tau \, ds \right\} \\
&\quad = \|f - g\| c_A^2 c_B^2 c_R \sup_{t_b \leq t \leq t_f} \left\{ (t_f - t_b) + \frac{c_Q}{2} \left((t_f - t_b)^2 - (t - t_b)^2 \right) \right\} \\
&\leq \|f - g\| c_A^2 c_B^2 c_{R^{-1}} (t_f - t_b)(1 + \frac{c_Q}{2}(t_f - t_b)) \,. \qquad (9.48b)
\end{aligned}
$$

Thus, $\mathcal{T}$ is a contraction if

$$
c_B^2 < \frac{1}{c_A^2 c_{R^{-1}} (t_f - t_b) \left(1 + \frac{c_Q}{2}(t_f - t_b)\right)} \qquad (9.48c)
$$

and therefore

$$
c_B < \frac{1}{c_A \sqrt{c_{R^{-1}} (t_f - t_b) \left(1 + \frac{c_Q}{2}(t_f - t_b)\right)}} \,. \qquad (9.48d)
$$

In order to get a condition that is independent of t_b, we take the worst case $t_b = t_0$, hence,

$$
c_B < \frac{1}{c_A \sqrt{c_{R^{-1}} (t_f - t_0) \left(1 + \frac{(t_f - t_0) c_Q}{2}\right)}} \,. \qquad (9.48e)
$$

□

Remark 9.4 Condition (9.48e) holds if the matrix Γ in (9.1c) has a sufficiently small Frobenius norm. Indeed, according to (9.2b) we have

$$B = \begin{pmatrix} 0 \\ M^{-1}\Gamma \end{pmatrix}$$

and therefore

$$c_B = \|B\|_F = \|M^{-1}\Gamma\|_F \leq \|M^{-1}\|_F \cdot \|\Gamma\|_F .$$

Having $\overline{y}^{(b)}(t)$, according to (9.44a) a stochastic optimal open-loop control $u^*(t) = u^*(t; t_b, \mathcal{I}_{t_b})$, $t_b \leq t \leq t_f$, reads

$$u^*(t; t_b, \mathcal{I}_{t_b}) = -R^{-1} B^T \overline{y(t)}^{(b)}. \tag{9.49a}$$

Moreover,

$$\varphi(t_b, \mathcal{I}_{t_b}) := u^*(t_b), t_0 \leq t_b \leq t_f \tag{9.49b}$$

is then a stochastic optimal open-loop feedback control law.

Remark 9.5 Putting $Q = 0$ in (9.40), we again obtain the stochastic optimal open-loop feedback control law (9.26) in Sect. 9.6.1.1.

9.8 Stochastic Weight Matrix $Q = Q(t, \omega)$

In the following we consider the case that, cf. (3.6h,i), the weight matrix for the evaluation of the displacements $z = z(t, \omega)$ is stochastic and may depend also on time t, $t_0 \leq t \leq t_f$. In order to take into account especially the size of the additive disturbance term $b = b(t, \omega)$, cf. (9.7b), in the following we consider the stochastic weight matrix

$$Q(t, \omega) := \|b(t, \omega)\|^2 Q, \tag{9.50a}$$

where Q is again a positive (semi) definite $2m \times 2m$ matrix, and $\|\cdot\|$ denotes the Euclidian norm.

According to (9.13c), (9.13d), for the adjoint variable $y = y(t, \omega)$ we then have the system of differential equations

$$\begin{aligned} \dot{y}(t, \omega) &= -A^T y(t, \omega) - \beta(t, \omega) Q z(t, \omega) \\ y(t_f, \omega) &= \nabla G(t_f, \omega, z(t_f, \omega)), \end{aligned}$$

where the stochastic function $\beta = \beta(t, \omega)$ is defined by

$$\beta(t, \omega) := \|b(t, \omega)\|^2. \tag{9.50b}$$

Assuming that we have also the weighted terminal costs,

$$G(t_f, \omega, z(t_f, \omega)) := \frac{1}{2}\beta(t_f, \omega)\|z(t_f, \omega) - z_f(\omega)\|^2, \tag{9.50c}$$

for the adjoint variable $y = y(t, \omega)$, we have the boundary value problem

$$\dot{y}(t, \omega) = -A^T y(t, \omega) - \beta(t, \omega) Q z(t, \omega) \tag{9.51a}$$

$$\begin{aligned} y(t_f, \omega) &= \beta(t_f, \omega)\left(z(t_f, \omega) - z_f(\omega)\right) \\ &= \beta(t_f, \omega) z(t_f, \omega) - \beta(t_f, \omega) z_f(\omega). \end{aligned} \tag{9.51b}$$

Corresponding to (9.40), from (9.51a), (9.51b) we then get the solution

$$\begin{aligned} y(t, \omega) = &\int_t^{t_f} e^{A^T(s-t)} Q\beta(s, \omega) z(s, \omega)\, ds \\ &+ e^{A^T(t_f - t)}\left(\beta(t_f, \omega) z(t_f, \omega) - \beta(t_f, \omega) z_f(\omega)\right), \quad t_b \le t \le t_f. \end{aligned} \tag{9.52a}$$

Taking conditional expectations of (9.52a) with respect to $\mathfrak{A}_{t_b}$, corresponding to (9.41) we obtain

$$\begin{aligned} \overline{y}^{(b)}(t) = &e^{A^T(t_f - t)}\left(\overline{\beta(t_f) z(t_f)}^{(b)} - \overline{\beta(t_f) z_f}^{(b)}\right) \\ &+ \int_t^{t_f} e^{A^T(s-t)} Q\overline{\beta(s) z(s)}^{(b)}\, ds, \quad t \ge t_b. \end{aligned} \tag{9.52b}$$

Since the matrices A, B are assumed to be fixed, see (9.7b), from (9.8c) and (9.52b) we get

$$h(t) = E\left(B^T y(t, \omega) | \mathfrak{A}_{t_b}\right) = B^T \overline{y}^{(b)}(t), \quad t \ge t_b. \tag{9.53a}$$

Consequently, corresponding to (9.11c) and (9.12), with (9.53a) the optimal open-loop control $u^* = u^*(t)$ is given then by

$$u^*(t; \mathcal{I}_{t_b}) = R^{-1}\left(-h(t)\right). \tag{9.53b}$$

Moreover, the weighted conditional mean trajectory

$$t \to \overline{\beta(t) z(t)}^{(b)} = E\left(\beta(t, \omega) z(t, \omega) | \mathfrak{A}_{t_b}\right), \quad t \ge t_b \tag{9.54a}$$

is determined in the present open-loop feedback approach as follows. We first remember that the optimal trajectory is defined by the initial value problem (9.13a), (9.13b). Approximating the weighted conditional mean trajectory (9.54a) by

$$t \to E\left(\beta(t_b,\omega)z(t,\omega)|\mathfrak{A}_{t_b}\right), \quad t_b \le t \le t_f, \tag{9.54b}$$

we multiply (9.13a), (9.13b) by $\beta(t_b,\omega)$. Thus, the trajectory $t \to \beta(t_b,\omega)z(t,\omega)$, $t \ge t_b$, is the solution of the initial value problem

$$\begin{aligned}\frac{d}{dt}\beta(t_b,\omega)z(t,\omega) &= A\beta(t_b,\omega)z(t,\omega) - BR^{-1}B\beta(t_b,\omega)\overline{y}^{(b)}(t)\\ &\quad + \beta(t_b,\omega)b(t,\omega) \qquad (9.55a)\\ \beta(t_b,\omega)z(t_b,\omega) &= \beta(t_b,\omega)z_b\,. \qquad (9.55b)\end{aligned}$$

Taking conditional expectations of (9.55a), (9.55b) with respect to $\mathfrak{A}_{t_b}$, for the approximate weighted conditional mean trajectory (9.54b) we obtain the initial value problem

$$\begin{aligned}\frac{d}{dt}\overline{\beta(t_b)z(t)}^{(b)} &= A\overline{\beta(t_b)z(t)}^{(b)} - BR^{-1}B\overline{\beta}^{(b)}(t_b)\overline{y}^{(b)}(t)\\ &\quad + \overline{\beta(t_b)b(t)}^{(b)} \qquad (9.56a)\\ \overline{\beta(t_b)z(t_b)}^{(b)} &= \overline{\beta(t_b)z_b}^{(b)}, \qquad (9.56b)\end{aligned}$$

where $\overline{\beta}^{(b)}(t) := E\left(\beta(t,\omega)|\mathfrak{A}_{t_b}\right)$, $t \ge t_b$. Consequently, the approximate weighted conditional mean trajectory (9.54b) can be represented, cf. (9.43a), (9.43b), by

$$\begin{aligned}\overline{\beta(t_b)z(t)}^{(b)} &= e^{A(t-t_b)}\overline{\beta(t_b)z_b}^{(b)}\\ &\quad + \int_{t_b}^{t} e^{A(t-s)}\left(\overline{\beta(t_b)b(s)}^{(b)} - BR^{-1}B^T\overline{\beta}^{(b)}(t_b)\overline{y}^{(b)}(s)\right)ds,\\ t_b \le t \le t_f\,. \qquad (9.57)\end{aligned}$$

Obviously, a corresponding approximate representation for

$$t \to \overline{\beta(t_f)z(t_f)}^{(b)}$$

can be obtained, cf. (9.43b), by using (9.57) for $t = t_f$.

Inserting now (9.57) into (9.52b), corresponding to (9.44c) we find the following approximate fixed point condition for the conditional mean adjoint trajectory $t \mapsto \overline{y}^{(b)}(t)$, $t_b \le t \le t_f$, needed in the representation (9.53b) of the stochastic optimal open-loop control $u^* = u^*(t)$, $t_b \le t \le t_f$:

$$\begin{aligned}
\overline{y}^{(b)}(t) &= e^{A^T(t_f - t)} \left(\overline{\beta(t_f) z(t_f)}^{(b)} - \overline{\beta(t_f) z_f}^{(b)} \right) + \int_{t_b}^{t_f} e^{A^T(s-t)} Q \overline{\beta(s) z(s)}^{(b)} \, ds \\
&\approx e^{A^T(t_f - t)} \Bigg(e^{A(t_f - t_b)} \overline{\beta(t_b) z_b}^{(b)} - \overline{\beta(t_f) z_f}^{(b)} \\
&\qquad + \int_{t_b}^{t_f} e^{A(t_f - s)} \left(\overline{\beta(t_b) b(s)}^{(b)} - B R^{-1} B^T \overline{\beta}^{(b)}(t_b) \overline{y}^{(b)}(s) \right) ds \Bigg) \\
&\quad + \int_{t}^{t_f} e^{A^T(s-t)} Q \Bigg(e^{A(s - t_b)} \overline{\beta(t_b) z_b}^{(b)} \\
&\qquad + \int_{t_b}^{s} e^{A(s-\tau)} \left(\overline{\beta(t_b) b(\tau)}^{(b)} - B R^{-1} B^T \overline{\beta}^{(b)}(t_b) \overline{y}^{(b)}(\tau) \right) d\tau \Bigg) ds.
\end{aligned} \tag{9.58}$$

Corresponding to Theorem 9.3 we can also develop a condition that guarantees the existence and uniqueness of a solution $\overline{y}^b = \overline{y}^{(b)}(t)$ of equation (9.58).

Theorem 9.4 *In the space of continuous functions, Eq. (9.58) has a unique solution if*

$$c_B < \frac{1}{c_A \sqrt{c_{R^{-1}} \overline{\eta}^{(b)}(t_b)(t_f - t_0) \left(1 + \frac{(t_f - t_0) c_Q}{2} \right)}}. \tag{9.59}$$

Here again,

$$c_A := \sup_{t_b \le t \le s \le t_f} \| e^{A(t-s)} \|_F \quad c_B := \|B\|_F \quad c_{R^{-1}} := \|R^{-1}\|_F \quad c_Q := \|Q\|_F,$$

and the index F denotes the Frobenius-Norm.

According to (9.53a), (9.53b), the stochastic optimal open-loop control $u^*(t)$, $t_b \le t \le t_f$, can be obtained as follows:

Theorem 9.5 *With a solution $\overline{y}^{(b)}(t)$ of the fixed point condition (9.58), the stochastic optimal open-loop control $u^*(t)$, $t_b \le t \le t_f$, reads*

$$u^*(t; \mathcal{I}_{t_b}) = -R^{-1} B^T \overline{y}^{(b)}(t). \tag{9.60a}$$

Moreover,

$$\varphi(t_b, \mathcal{I}_{t_b}) := u^*(t_b; \mathcal{I}_{t_b}), \quad t_0 \le t_b \le t_f \tag{9.60b}$$

is then the stochastic optimal open-loop feedback control law.

9.9 Uniformly Bounded Sets of Controls D_t, $t_0 \le t \le t_f$

The above-shown Theorem 9.4 guaranteeing the existence of a solution of the fixed point condition (9.58) can be generalized considerably if we suppose that the sets D_t of feasible controls $u = u(t)$ are uniformly bounded with respect to the time t, $t_0 \le t \le t_f$. Hence, in the following we suppose again:

- time-independent and deterministic matrices of coefficients, hence

$$A(t,\omega) = A \qquad B(t,\omega) = B \tag{9.61a}$$

- quadratic cost functions,

$$\begin{aligned} C(u) &= \frac{1}{2}u^T Ru \qquad Q(z) = \frac{1}{2}z^T Qz \\ G(t_f,\omega,z(t_f,\omega)) &= \frac{1}{2}\left\|z(t_f,\omega) - z_f(\omega)\right\|_2^2 \end{aligned} \tag{9.61b}$$

- uniformly bounded sets of feasible controls, hence, we assume that there exists a constant $C_D \ge 0$ such that

$$\|u\|_2 \le c_D \quad \text{for all } u \in \bigcup_{t\in T} D_t. \tag{9.61c}$$

According to the above assumed deterministic coefficient matrices A, b, the H-minimal control depends only on the conditional expectation of the adjoint trajectory, hence,

$$\widetilde{u^*}(t) = \widetilde{u^*}(t, \overline{y(t)}^{(b)}). \tag{9.62}$$

Thus, the integral form of the related 2-point boundary value problem reads

$$z(\omega,t) = z_b + \int_{t_b}^{t} \left(Az(\omega,s) + b(s,\omega) + B\widetilde{u^*}(s, \overline{y(s)}^{(b)}) \right) ds \tag{9.63a}$$

$$y(\omega,t) = \left(z(t_f,\omega) - z_f(\omega)\right) + \int_{t}^{t_f} \left(A^T y(\omega,s) + Qz(s,\omega)\right) ds. \tag{9.63b}$$

Consequently, for the conditional expectations of the trajectories we get

$$\overline{z(t)}^{(b)} = \overline{z_b}^{(b)} + \int_{t_b}^{t} \left(A\overline{z(s)}^{(b)} + \overline{b(s)}^{(b)} + B\widetilde{u^*}(s, \overline{y(s)}^{(b)}) \right) ds \tag{9.64a}$$

$$\overline{y(t)}^{(b)} = \left(\overline{z(t_f)}^{(b)} - \overline{z_f}^{(b)} \right) + \int_{t}^{t_f} \left(A^T \overline{y(s)}^{(b)} + Q\overline{z(s)}^{(b)} \right) ds. \tag{9.64b}$$

Using the matrix exponential function with respect to A, we have

$$\overline{z(t)}^{(b)} = e^{A(t-t_b)}\overline{z_b}^{(b)} + \int_{t_b}^{t} e^{A(t-s)} \left(\overline{b(s)}^{(b)} + B\widetilde{u^*}(s, \overline{y(s)}^{(b)}) \right) ds \tag{9.65a}$$

$$\overline{y(t)}^{(b)} = e^{A^T(t_f-t)} \left(\overline{z(t_f)}^{(b)} - \overline{z_f}^{(b)} \right) + \int_{t}^{t_f} e^{A^T(s-t)} Q\overline{z(s)}^{(b)} ds. \tag{9.65b}$$

Putting (9.65a) into (9.65b), for $\overline{y(t)}^{(b)}$ we get then the following fixed point condition

$$\begin{aligned} \overline{y(t)}^{(b)} &= e^{A^T(t_f-t)} \left(\overline{z(t_f)}^{(b)} - \overline{z_f}^{(b)} \right) \\ &+ \int_{t}^{t_f} e^{A^T(s-t)} Q \left(e^{A(s-t_b)}\overline{z_b}^{(b)} + \int_{t_b}^{s} e^{A(s-\tau)} \left(\overline{b(\tau)}^{(b)} + B\widetilde{u^*}(\tau, \overline{y(\tau)}^{(b)}) \right) d\tau \right) ds \end{aligned} \tag{9.66}$$

For the consideration of the existence of a solution of the above fixed point equation (9.66) we need several auxiliary tools. According to the assumption (9.61c), next we have the following lemma:

Lemma 9.4 *There exist constants $c_z, c_G > 0$ such that*

$$\|\overline{z(f(\cdot), t)}^{(b)}\| \le c_z \qquad \text{and} \qquad \|\overline{\nabla_z G(z(f(\cdot), t)}^{(b)}\| \le c_G \tag{9.67a}$$

for each time $t, t_b \le t \le t_f$, and all $f \in C(T; \mathbb{R}^m)$, where

$$\overline{z(f(\cdot), t)}^{(b)} := e^{A(t-t_b)}\overline{z_b}^{(b)} + \int_{t_b}^{t} e^{A(t-s)} \left(\overline{b(s)}^{(b)} + B\widetilde{u^*}(f(s), s) \right) ds \tag{9.67b}$$

Proof With $c_A := e^{\|A\|_F(t_f-t_b)}$, $c_B := \|B\|_F$ and $c_{\overline{b}^{(b)}} := \|\overline{b(\cdot)}^{(b)}\|_\infty$ the following inequalities hold:

$$\|\overline{z(f(\cdot),t)}^{(b)}\|_2 \le c_A\left(\|\overline{z_b}^{(b)}\|_2 + \left(c_{\overline{b}^{(b)}} + c_B c_D\right)(t_f - t_b)\right) \le c_z \tag{9.68a}$$

$$\|\overline{\nabla_z G(z(f(\cdot),t)}^{(b)}\|_2 = \|\overline{z(f(\cdot),t_f)}^{(b)} - \overline{z_f}^{(b)}\| \le c_z + \|\overline{z_f}^{(b)}\|_2 \le c_G, \tag{9.68b}$$

where c_z, c_G are arbitrary upper bounds of the corresponding left quantities. □

In the next lemma the operator defined by the right-hand side of (9.66) is studied:

Lemma 9.5 *Let denote again $\mathcal{X} := C(T;\mathbb{R}^m)$ the space of continuous functions f on T equipped with the supremum norm. If $\tilde{\mathcal{T}} : \mathcal{X} \to \mathcal{X}$ denotes the operator defined by*

$$\begin{aligned}(\tilde{\mathcal{T}} f)(t) &:= e^{A^T(t_f-t)}\left(\overline{z(t_f)}^{(b)} - \overline{z_f}^{(b)}\right)\\ &+ \int_t^{t_f} e^{A^T(s-t)} Q\left(e^{A(s-t_b)}\overline{z_b}^{(b)} + \int_{t_b}^{s} e^{A(s-\tau)}\left(\overline{b(\tau)}^{(b)} + B\widetilde{u^*}(\tau, f(\tau))\right) d\tau\right) ds,\end{aligned} \tag{9.69}$$

then the image of $\tilde{\mathcal{T}}$ is relative compact.

Proof Let $c_Q := \|Q\|_F$. We have to show that $\tilde{\mathcal{T}}(\mathcal{X})$ is bounded and equicontinuous.

- $\tilde{\mathcal{T}}(\mathcal{X})$ is bounded:

$$\begin{aligned}&\left\| e^{A^T(t_f-t)}\left(\overline{z(t_f)}^{(b)} - \overline{z_f}^{(b)}\right)\right.\\ &\left. + \int_t^{t_f} e^{A^T(s-t)} Q\left(e^{A(s-t_b)}\overline{z_b}^{(b)} + \int_{t_b}^{s} e^{A(s-\tau)}\left(\overline{b(\tau)}^{(b)} + B\widetilde{u^*}(\tau, f(\tau))\right) d\tau\right) ds\right\|\end{aligned}$$

$$\le c_A c_G + c_A c_Q\left(c_A\|\overline{z_b}^{(b)}\|_2(t_f - t_b) + (c_A c_{\overline{b}^{(b)}} + c_B c_D)\frac{t_f^2 - t_b^2}{2}\right) \tag{9.70}$$

- $\tilde{\mathcal{T}}(\mathcal{X})$ is equicontinuous.
  We have to show that for each $\epsilon > 0$ there exists a $\delta > 0$ such that, independent of the mapped function f, the following inequality holds:

$$|t - s| < \delta \quad \Rightarrow \quad \|\tilde{\mathcal{T}} f(t) - \tilde{\mathcal{T}} f(s)\|_2 \le \epsilon.$$

Defining the function

$$\varrho(t) = e^{A(t-t_b)} z_b + \int_{t_b}^{t} e^{A(t-\mu)} \overline{b(\mu)}^{(b)} \, d\mu, \tag{9.71}$$

the following inequalities hold:

$$\begin{aligned}
&\left\| \tilde{\mathcal{T}} f(t) - \tilde{\mathcal{T}} f(s) \right\|_2 \\
&= \left\| e^{A^T (t_f - t)} \left(\overline{z(t_f)}^{(b)} - \overline{z_f}^{(b)} \right) - e^{A^T (t_f - s)} \left(\overline{z(t_f)}^{(b)} - \overline{z_f}^{(b)} \right) \right. \\
&\quad + \int_t^{t_f} e^{A^T (\tau - t)} Q \left(e^{A(\tau - t_b)} z_b + \int_{t_b}^{\tau} e^{A(\tau - \mu)} \left(\overline{b(\mu)}^{(b)} + B \widetilde{u^*}(\mu, f(\mu)) \right) d\mu \right) d\tau \\
&\quad \left. - \int_s^{t_f} e^{A^T (\tau - s)} Q \left(e^{A(\tau - t_b)} z_b + \int_{t_b}^{\tau} e^{A(\tau - \mu)} \left(\overline{b(\mu)}^{(b)} + B \widetilde{u^*}(\mu, f(\mu)) \right) d\mu \right) d\tau \right\|.
\end{aligned} \tag{9.72a}$$

From Eq. (9.72a) we get then

$$\begin{aligned}
&\left\| \tilde{\mathcal{T}} f(t) - \tilde{\mathcal{T}} f(s) \right\|_2 \\
&= \left\| \left(e^{A^T (t_f - t)} - e^{A^T (t_f - s)} \right) \left(\overline{z(t_f)}^{(b)} - \overline{z_f}^{(b)} \right) \right. \\
&\quad + \int_t^{t_f} e^{A^T (\tau - t)} Q \left(\varrho(\tau) + \int_{t_b}^{\tau} B \widetilde{u^*}(\mu, f(\mu)) \, d\mu \right) d\tau \\
&\quad \left. - \int_s^{t_f} e^{A^T (\tau - s)} Q \left(\varrho(\tau) + \int_{t_b}^{\tau} B \widetilde{u^*}(\mu, f(\mu)) \, d\mu \right) d\tau \right\|_2 \\
&\le \left\| e^{A^T (t_f - t)} - e^{A^T (t_f - s)} \right\|_2 c_G \\
&\quad + \left\| \left(e^{A^T (-t)} - e^{A^T (-s)} \right) \int_t^{t_f} e^{A^T \tau} Q \left(\varrho(\tau) + \int_{t_b}^{\tau} B \widetilde{u^*}(\mu, f(\mu)) \, d\mu \right) d\tau \right. \\
&\quad \left. - e^{A^T (-s)} \int_s^{t} e^{A^T \tau} Q \left(\varrho(\tau) + \int_{t_b}^{\tau} B \widetilde{u^*}(\mu, f(\mu)) \, d\mu \right) d\tau \right\|_2 \\
&\le \left\| e^{A^T (t_f - t)} - e^{A^T (t_f - s)} \right\|_2 c_G \\
&\quad + \left\| e^{A^T (-t)} - e^{A^T (-s)} \right\| e^{\|A\|_F t_f} c_Q (c_\varrho + c_B c_D (t_f - t_b)) (t_f - t_b) \\
&\quad + c_A c_Q (c_\varrho + c_B c_D (t_f - t_b)) |t - s|.
\end{aligned} \tag{9.72b}$$

Obviously, the final expression in (9.72b) is independent of $f(\cdot)$. Hence, due to the continuity of the matrix exponential function and the function $\varrho(\cdot)$, the assertion follows. □

From the above Lemma 9.5 we now obtain this result:

Theorem 9.6 *The fixed point Eq. (9.66) has a continuous, bounded solution.*

Proof Define again$\mathcal{X} := C(T; \mathbb{R}^m)$ and consider the set $\mathcal{M} \subset \mathcal{X}$

$$\mathcal{M} := \left\{ f(\cdot) \in \mathcal{X} \mid \sup_{t \in T} \| f(t) \|_2 \leq C \right\}, \tag{9.73a}$$

where

$$C := c_A c_G + c_A c_Q \left(c_A \| z_b \|_2 (t_f - t_b) + (c_A c_{\overline{b}^{(b)}} + c_B c_D) \frac{t_f^2 - t_b^2}{2} \right). \tag{9.73b}$$

Moreover, let $\mathcal{T}$ denote the restriction of $\tilde{\mathcal{T}}$ to $\mathcal{M}$, hence,

$$\mathcal{T} : \mathcal{M} \to \mathcal{M}, \quad f \mapsto \tilde{\mathcal{T}} f. \tag{9.74}$$

Obviously, the operator $\mathcal{T}$ is continuous and, according to Lemma 9.5, the image of $\mathcal{T}$ is relative compact. Moreover, the set $\mathcal{M}$ is closed and convex. Hence, according to the fixed point theorem of Schauder, $\mathcal{T}$ has a fixed point in $\mathcal{M}$. □

9.10 Approximate Solution of the Two-Point Boundary Value Problem (BVP)

According to the previous sections, the remaining problem is then to solve the fixed point Eq. (9.44c) or (9.58). In the first case, the corresponding equation reads

$$\begin{aligned} \overline{y}^{(b)}(t) = e^{A^T(t_f - t)} G_f \Bigg(& e^{A(t_f - t_b)} \overline{z_b}^{(b)} - \overline{z_f}^{(b)} \\ & + \int_{t_b}^{t_f} e^{A(t_f - s)} \left(\overline{b(s)}^{(b)} - B R^{-1} B^T \overline{y}^{(b)}(s) \right) ds \Bigg) \\ & + \int_{t}^{t_f} e^{A^T(s-t)} Q \Bigg(e^{A(s - t_b)} \overline{z_b}^{(b)} \\ & + \int_{t_b}^{s} e^{A(s-\tau)} \left(\overline{b(\tau)}^{(b)} - B R^{-1} B^T \overline{y}^{(b)}(\tau) \right) d\tau \Bigg) ds \,. \end{aligned} \tag{9.75}$$

Based on the present stochastic open-loop feedback control approach, we present the following approximation method:

9.10.1 Approximate Solution of the Fixed Point Eq. (9.75)

Step I According to the equations (9.12) and (9.44a) of the stochastic optimal $OLFC$, for each remaining time interval $[t_b, t_f]$ the value of the stochastic optimal open-loop control $u^* = u^*(t; t_b, \mathcal{I}_{t_b}), t \le t_b$, is needed at the left time point t_b only.

Thus, putting first $t = t_b$ in (9.75), we get

$$\begin{aligned}
\overline{y}^{(b)}(t_b) &= e^{A^T(t_f - t_b)} G_f e^{A(t_f - t_b)} \overline{z_b}^{(b)} - e^{A^T(t_f - t_b)} G_f \overline{z_f}^{(b)} \\
&\quad + e^{A^T(t_f - t_b)} G_f \int_{t_b}^{t_f} e^{A(t_f - s)} \overline{b(s)}^{(b)} \, ds \\
&\quad - e^{A^T(t_f - t_b)} G_f \int_{t_b}^{t_f} e^{A(t_f - s)} B R^{-1} B^T \overline{y}^{(b)}(s) \, ds \\
&\quad + \int_{t_b}^{t_f} e^{A^T(s - t_b)} Q e^{A(s - t_b)} \, ds \overline{z_b}^{(b)} \\
&\quad + \int_{t_b}^{t_f} e^{A^T(s - t_b)} Q \Big(\int_{t_b}^{s} e^{A(s - \tau)} \overline{b(\tau)}^{(b)} \, d\tau \Big) ds \\
&\quad - \int_{t_b}^{t_f} e^{A^T(s - t_b)} Q \Big(\int_{t_b}^{s} e^{A(s - \tau)} B R^{-1} B^T \overline{y}^{(b)}(\tau) \, d\tau \Big) ds \, . \qquad (9.76a)
\end{aligned}$$

Step II Due to the representation (9.44a) of the stochastic optimal open-loop control u^* and the stochastic OLF construction principle (3.10d,e), the value of the conditional mean adjoint variable $\overline{y}^{(b)}(t)$ is needed at the left boundary point $t = t_b$ only. Consequently, $\overline{y}^b = \overline{y}^{(b)}(s)$ is approximated on $[t_b, t_f]$ by the constant function

$$\overline{y}^{(b)}(s) \approx \overline{y}^{(b)}(t_b), t_b \le s \le t_f . \qquad (9.76b)$$

In addition, the related matrix exponential function $s \to e^{A(t_f - s)}$ is approximated on $[t_b, t_f]$ in the same way.

This approach is justified especially if one works with a *receding time horizon* or *moving time horizon*

$$t_f := t_b + \Delta$$

with a short *prediction time horizon* Δ.

$$\begin{aligned}\overline{y}^{(b)}(t_b) \approx & e^{A^T(t_f-t_b)}G_f e^{A(t_f-t_b)}\overline{z_b}^{(b)} - e^{A^T(t_f-t_b)}G_f\overline{z_f}^{(b)} \\ & + e^{A^T(t_f-t_b)}G_f\int_{t_b}^{t_f} e^{A(t_f-s)}\overline{b(s)}^{(b)}\,ds \\ & - (t_f-t_b)e^{A^T(t_f-t_b)}G_f e^{A(t_f-t_b)}BR^{-1}B^T\overline{y}^{(b)}(t_b) \\ & + \int_{t_b}^{t_f} e^{A^T(s-t_b)}Qe^{A(s-t_b)}ds\,\overline{z_b}^{(b)} \\ & + \int_{t_b}^{t_f} e^{A^T(s-t_b)}Q\Big(\int_{t_b}^{s} e^{A(s-\tau)}\overline{b(\tau)}^{(b)}\,d\tau\Big)\,ds \\ & - \int_{t_b}^{t_f}(s-t_b)e^{A^T(s-t_b)}Qe^{A(s-t_b)}\,ds\,BR^{-1}B^T\overline{y}^{(b)}(t_b). \end{aligned} \tag{9.76c}$$

Step III Rearranging terms, (9.76c) yields a system of linear equations for $\overline{y}^{(b)}(t_b)$:

$$\begin{aligned}\overline{y}^{(b)}(t_b) \approx & A_0((t_b,t_f,G_f,Q)\overline{z_b}^{(b)} - e^{A^T(t_f-t_b)}G_f\overline{z_f}^{(b)} \\ & + A_1(t_b,t_f,G_f,Q)\cdot\overline{b}^{(b)}_{[t_b,t_f]}(\cdot) \\ & - A_{23}(t_b,t_f,G_f,Q)BR^{-1}B^T\overline{y}^{(b)}(t_b), \end{aligned} \tag{9.76d}$$

where the matrices, linear operator and function, resp., A_0, A_1, A_{23}, $\overline{b}^{(b)}_{[t_b,t_f]}$ can be easily read from relation (9.76c). Consequently, (9.76d) yields

$$\begin{aligned}\Big(I + A_{23}(t_b,t_f,G_f,Q)BR^{-1}B^T\Big)\overline{y}^{(b)}(t_b) \approx & A_0((t_b,t_f,G_f,Q)\overline{z_b}^{(b)} \\ & -e^{A^T(t_f-t_b)}G_f\overline{z_f}^{(b)} + A_1(t_b,t_f,G_f,Q)\cdot\overline{b}^{(b)}_{[t_b,t_f]}(\cdot). \end{aligned} \tag{9.76e}$$

For the matrix occurring in (9.76e) we have this result:

Lemma 9.6 *The matrix* $I + A_{23}(t_b, t_f, G_f, Q)BR^{-1}B^T$ *is regular.*

Proof According to (9.76c), (9.76d) we have

$$A_{23}(t_b, t_f, G_f, Q) = (t_f - t_b)e^{A^T(t_f - t_b)}G_f e^{A(t_f - t_b)} + \int_{t_b}^{t_f} (s - t_b)e^{A^T(s-t_b)}Qe^{A(s-t_b)}\,ds\ .$$

Hence, $A_{23} = A_{23}(t_b, t_f, G_f, Q)$ is a positive definite matrix. Moreover, $U := BR^{-1}B^T$ is at least positive semidefinite. Consider now the equation $(I + A_{23}U)w = 0$. We get $A_{23}Uw = -w$ and therefore $Uw = -A_{23}^{-1}w$, hence, $(U + A_{23}^{-1})w = 0$. However, since the matrix $U + A_{23}^{-1}$ is positive definite, we have $w = 0$, which proves now the assertion. □

The above lemma and (9.76e) yields now

$$\overline{y}^{(b)}(t_b) \approx \Big(I + A_{23}(t_b, t_f, G_f, Q)BR^{-1}B^T\Big)^{-1}\Big(A_0((t_b, t_f, G_f, Q)\overline{z_b}^{(b)} - e^{A^T(t_f - t_b)}G_f\overline{z_f}^{(b)} + A_1(t_b, t_f, G_f, Q)\cdot \overline{b}^{(b)}_{[t_b,t_f]}(\cdot)\Big). \quad (9.76f)$$

According to (9.75) and (3.10d,e), the stochastic optimal open-loop feedback control $\varphi^* = \varphi^*(t_b, \mathcal{I}_{t_b})$ at $t = t_b$ is obtained as follows:

Theorem 9.7 *With the approximative solution* $\overline{y}^{(b)}(t_b)$ *of the fixed point condition (9.75) at* $t = t_b$*, represented by (9.76f), the stochastic optimal open-loop feedback control law at* $t = t_b$ *is given by*

$$\varphi^*(t_b, \mathcal{I}_{t_b}) := -R^{-1}B^T\overline{y}^{(b)}(t_b). \quad (9.76g)$$

Moreover, the whole approximate stochastic optimal open-loop feedback control law $\varphi^* = \varphi^*(t, \mathcal{I}_t)$ *is obtained from (9.76g) by replacing* $t_b \to t$ *for arbitrary* t, $t_0 \le t \le t_f$.

9.11 Example

We consider the structure according to Fig.9.2, see [3], where we want to control the supplementary active system while minimizing the expected total costs for the control and the terminal costs.

The behavior of the vector of displacements $q(t, \omega)$ can be described by a system of differential equations of second order:

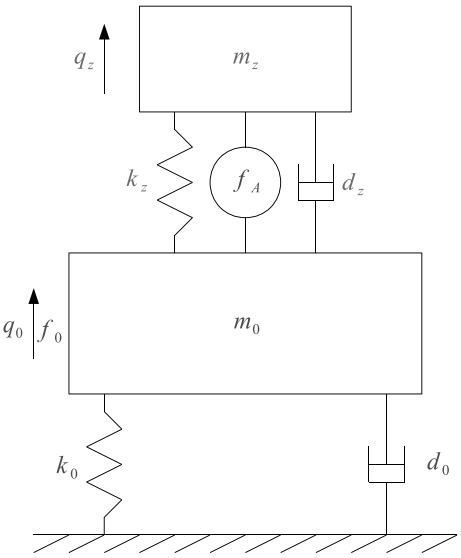


Fig. 9.2 Principle of active structural control

$$M\begin{pmatrix}\ddot{q}_0(t,\omega)\\ \ddot{q}_z(t,\omega)\end{pmatrix}+D\begin{pmatrix}\dot{q}_0(t,\omega,t\\ \dot{q}_z(t,\omega)\end{pmatrix}+K\begin{pmatrix}q_0(t,\omega)\\ q_z(t,\omega)\end{pmatrix}=f_0(t,\omega)+f_a(t) \tag{9.77}$$

with

$$M=\begin{pmatrix}m_0 & 0\\ 0 & m_z\end{pmatrix} \quad \text{mass matrix} \tag{9.78a}$$

$$D=\begin{pmatrix}d_0+d_z & -d_z\\ -d_z & d_z\end{pmatrix} \quad \text{damping matrix} \tag{9.78b}$$

$$K=\begin{pmatrix}k_0+k_z & -k_z\\ -k_z & k_z\end{pmatrix} \quad \text{stiffness matrix} \tag{9.78c}$$

$$f_a(t)=\begin{pmatrix}-1\\ +1\end{pmatrix}u(t) \quad \text{actuator force} \tag{9.78d}$$

$$f_0(t,\omega)=\begin{pmatrix}f_{01}(t,\omega)\\ 0\end{pmatrix} \quad \text{applied load .} \tag{9.78e}$$

Here we have $n=1$, i.e., $u(\cdot)\in C(T,\mathbb{R})$, and the weight matrix R becomes a positive real number.

To represent the equation of motion (9.77) as a first-order differential equation we set

$$z(t,\omega):=(q(t,\omega),\dot{q}(t,\omega))^T=\begin{pmatrix}q_0(t,\omega)\\ q_z(t,\omega)\\ \dot{q}_0(t,\omega)\\ \dot{q}_z(t,\omega)\end{pmatrix}.$$

This yields the dynamical equation

$$\dot{z}(t,\omega) = \begin{pmatrix} \underline{\mathbf{0}} & I_2 \\ -M^{-1}K & -M^{-1}D \end{pmatrix} z(t,\omega) + \begin{pmatrix} \underline{\mathbf{0}} \\ M^{-1} f_a(s) \end{pmatrix} + \begin{pmatrix} \underline{\mathbf{0}} \\ M^{-1} f_0(s,\omega) \end{pmatrix} =$$

$$= \underbrace{\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{k_0+k_z}{m_0} & \frac{k_z}{m_0} & -\frac{d_0+d_z}{m_0} & \frac{d_z}{m_0} \\ \frac{k_z}{m_z} & -\frac{k_z}{m_z} & \frac{d_z}{m_z} & -\frac{d_z}{m_z} \end{pmatrix}}_{:=A} z(t,\omega) + \underbrace{\begin{pmatrix} 0 \\ 0 \\ -\frac{1}{m_0} \\ \frac{1}{m_z} \end{pmatrix}}_{:=B} u(s) + \underbrace{\begin{pmatrix} 0 \\ 0 \\ \frac{f_0(s,\omega)}{m_0} \\ 0 \end{pmatrix}}_{:=b(t,\omega)}, \tag{9.79}$$

where I_p denotes the $p \times p$ identity matrix. Furthermore, we have the optimal control problem under stochastic uncertainty:

$$\min \quad F(u(\cdot)) := E\frac{1}{2}\left(\int_{t_b}^{t_f} R\,(u(s))^2 \; ds + z(t_f,\omega)^T G z(t_f,\omega) \Big| \, \mathfrak{A}_{t_b} \right) \tag{9.80a}$$

$$s.t. \quad z(t,\omega) = z_b + \int_{t_b}^{t} \big(Az(s,\omega) + Bu(s) + b(s,\omega) \big) ds \tag{9.80b}$$

$$u(\cdot) \in C(T,\mathbb{R}). \tag{9.80c}$$

Note that this problem is of the "Minimum-Energy Control"-type, if we apply no extra costs for the displacements, i.e., $Q \equiv 0$.

The two-point-boundary problem to be solved reads then, cf. (9.13a)–(9.13d),

$$\dot{z}(t,\omega) = Az(t,\omega) - \frac{1}{R} B B^T \overline{y(t)}^{(b)} + b(\omega,t) \tag{9.81a}$$

$$\dot{y}(t,\omega) = -A^T y(t,\omega) \tag{9.81b}$$

$$z(t_b,\omega) = z_b \tag{9.81c}$$

$$y(t_f,\omega) = Gz(t_f,\omega). \tag{9.81d}$$

Hence, the solution of (9.81a)–(9.81d), i.e., the optimal trajectories, reads, cf. (9.14a), (9.37a),

$$y(t,\omega) = e^{A^T(t_f-t)} G z(t_f,\omega) \tag{9.82a}$$

$$z(t,\omega) = e^{A(t-t_b)} z_b + \int_{t_b}^{t} e^{A(t-s)} \Big(b(s,\omega) - \frac{1}{R} B B^T e^{A^T(t_f-s)} G \overline{z(t_f)}^{(b)} \Big) ds. \tag{9.82b}$$

Finally, we get the optimal control, see (9.38c) and (9.39) :

$$u^*(t) = -\frac{1}{R} B^T e^{A^T(t_f - t)} \left(I_4 + GU\right)^{-1} G e^{At_f} \left(e^{-At_b} z_b + \int_{t_b}^{t_f} e^{-As} \overline{b(s)}^{(b)} \, ds \right) \tag{9.83}$$

with

$$U = \frac{1}{R} \int_{t_b}^{t_f} e^{A(t_f - s)} B B^T e^{A^T(t_f - s)} \, ds. \tag{9.84}$$

References

1. Allgöwer, F.E.A. (ed.): Nonlinear Model Predictive Control. Birkhäuser Verlag, Basel (2000)
2. Aoki, M.: Optimization of Stochastic Systems - Topics in Discrete-Time Systems. Academic, New York (1967)
3. Block, C.: Aktive Minderung personeninduzierter Schwingungen an weit gespannten Strukturen im Bauwesen. No. 336 in Fortschrittberichte VDI, Reihe 11, Schwingungstechnik. VDI-Verlag GmbH, Düsseldorf (2008)
4. Dullerud, G., Paganini, F.: A Course in Robust Control Theory. Springer, New York (2000)
5. Ku, R., Athans, M.: On the adaptive control of linear systems using the open-loop feedback optimal approach. IEEE Trans. Autom. Control **18**, 489–493 (1973)
6. Marti, K.: Stochastic optimization methods in robust adaptive control of robots. In: Groetschel, M.E.A. (ed.) Online Optimization of Large Scale Systems, pp. 545–577. Springer, Berlin (2001)
7. Marti, K.: Adaptive Optimal Stochastic Trajectory Planning and Control (AOSTPC) for Robots, pp. 155–206. Springer, Berlin (2004)
8. Marti, K.: Approximate solutions of stochastic control problems by means of convex approximations. In: Topping, B. et al. (eds.) Proceedings of the 9th International Conference on Computational Structures Technology (CST08). Civil-Comp Press, Stirlingshire (2008)
9. Marti, K.: Stochastic nonlinear model predictive control (snmpc). In: 79th Annual Meeting of the International Association of Applied Mathematics and Mechanics (GAMM), Bremen 2008, PAMM, vol. 8, Issue 1, pp. 10775–10776. Wiley-VCH (2008)
10. Marti, K.: Stochastic Optimization Methods, 2nd edn. Springer, Berlin (2008). https://doi.org/10.1007/978-3-540-79458-5
11. Marti, K.: Continuous-time control under stochastic uncertainty. In: Cochran, J.E.A. (ed.) Wiley Encyclopedia of Operations Research and Management Science (EORMS). Wiley, Hoboken (2010). https://doi.org/10.1002/9780470400531.eorms0839
12. Marti, K.: Optimal control of dynamical systems and structures under stochastic uncertainty: stochastic optimal feedback control. Adv. Engin. Softw. (AES) **46**, 43–62 (2012). https://doi.org/10.1016/j.advengsoft.2010.09.008
13. Marti, K.: Stochastic optimal structural control: stochastic optimal open-loop feedback control. Adv. Eng. Softw. **44**(1), 26–34 (2012). https://doi.org/10.1016/j.advengsoft.2011.05.040. CIVIL-COMP
14. Nagarajaiah, S., Narasimhan, S.: Optimal control of structures. In: Arora, J. (ed.) Optimization of Structural and Mechanical Systems, pp. 221–244. World Scientific, New Jersey (2007)
15. Ostrowski, A.: Über Eigenwerte von Produkten Hermitescher Matrizen. Abh. Math. Semin. Univ. Hambg. **23**, 60–68 (1959)

16. Richalet, J., et al.: Model predictive heuristic control: applications to industrial processes. Automatica **14**, 413–428 (1978). https://doi.org/10.1016/0005-1098(78)90001-8
17. Schacher, M.: Stochastisch optimale Regelung von Robotern. No. 1200 in Fortschritt-Berichte VDI, Reihe 8, Mess-, Steuerungs- und Regelungstechnik. VDI Verlag GmbH, Düsseldorf (2011)
18. Soong, T.: Active structural control in civil engineering. Eng. Struct. **10**, 74–84 (1988)
19. Soong, T.: Active Structural Control: Theory and Practice. Wiley, New York (1990)
20. Soong, T., Costantinou, M.: Passive and Active Structural Vibration Control in Civil Engineering, CISM Courses and Lectures, vol. 345. Springer, Wien (1994)
21. Spencer, B., Nagarajaiah, S.: State of the art of structural control. J. Struct. Eng. **129**(7), 845–856 (2003). https://doi.org/10.1061/(ASCE)0733-9445(2003)129:7(845)
22. Yang, J., Soong, T.: Recent advances in active control of civil engineering structures. Probab. Eng. Mech. **3**(4), 179–188 (1988). https://doi.org/10.1016/0266-8920(88)90010-0

Chapter 10
Adaptive Optimal Stochastic Trajectory Planning and Control (AOSTPC)

Abstract Adaptive Optimal Stochastic Trajectory Planning and Control (*AOSTPC*) are considered in this chapter: In optimal control of dynamic systems the standard procedure is to determine first offline an optimal open-loop control, using some nominal or estimated values of the model parameters, and to correct then the resulting deviation of the actual trajectory or system performance from the prescribed trajectory (prescribed system performance) by online measurement and control actions. However, online measurement and control actions are very expensive and time consuming. By adaptive optimal stochastic trajectory planning and control (AOSTPC), based on stochastic optimization methods, the available a priori and statistical information about the unknown model parameters is incorporating into the optimal control design. Consequently, the mean absolute deviation between the actual and prescribed trajectory can be reduced considerably, and robust controls are obtained. Using only some necessary stability conditions, by means of stochastic optimization methods also sufficient stability properties of the corresponding feedforward, feedback (PD-, PID-) controls, resp., are obtained. Moreover, analytical estimates are given for the reduction of the tracking error, hence, for the reduction of the online correction expenses by applying (AOSTPC).

10.1 Introduction

An industrial, service, or field robot is modeled mathematically by its **dynamic equation**, being a system of second-order differential equations for the robot or configuration coordinates $q = (q_1, \ldots, q_n)^T$ (rotation angles in case of revolute links, length of translations in case of prismatic links), and the **kinematic equation**, relating the space $\{q\}$ of robot coordinates to the work space $\{x\}$ of the robot. Thereby one meets [4, 34, 42, 45, 47, 49] several model parameters, such as length of links, $l_i(m)$, location of center of gravity of links, $l_{ci}(m)$, mass of links, $m_i(kg)$, payload (N), moments of inertia about centroid, $I_i(kgm^2)$, (Coulomb-) friction coefficients, $R_{ij0}(N)$, etc. Let p_D, p_K denote the vector of model parameters contained in the dynamic, kinematic equation, respectively. A further vector p_C of model parameters occurs in the formulation of several constraints, especially initial and terminal

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9_10

conditions, control and state constraints of the robot, as, e.g., maximum, minimum torques or forces in the links, bounds for the position, maximum joint, path velocities. Moreover, certain parameters p_J, e.g., cost factors, may occur also in the objective (performance, goal) functional J.

Due to stochastic variations of the material, manufacturing errors, measurement (identification) errors, stochastic variations of the workspace environment, as, e.g., stochastic uncertainty of the payload, randomly changing obstacles, errors in the selection of appropriate bounds for the moments, forces, resp., in the links, for the position and path velocities, errors in the selection of random cost factors, modeling errors, disturbances, etc., the total vector

$$p = \begin{pmatrix} p_D \\ p_K \\ p_C \\ p_J \end{pmatrix} \tag{10.1a}$$

of model parameters is not a given fixed quantity. The vector p must be represented therefore by a random vector

$$p = p(\omega), \quad \omega \in (\Omega, \mathcal{A}, P) \tag{10.1b}$$

on a certain probability space $(\Omega, \mathcal{A}, P)$, see [3, 14, 32, 45, 50].

Having to control a robotic or more general dynamical system, the control law $u = u(t)$, is represented usually by the sum

$$u(t) := u^{(0)}(t) + \Delta u(t), \quad t_0 \le t \le t_f, \tag{10.2}$$

of a feedforward control (open-loop-control) $u_0(t)$, $t_0 \le t \le t_f$, and an online local control correction (feedback control) $\Delta u(t)$.

In actual engineering practice [19, 33, 35, 51], the feedforward control $u^{(0)}(t)$ is determined offline based on a certain reference trajectory $q^{(0)}(t)$, $t_0 \le t \le t_f$, in configuration space, where the unknown parameter vector p is replaced by a certain vector $p^{(0)}$ of nominal parameter values, as, e.g., the expectation $p^{(0)} := \overline{p} = Ep(\omega)$. The increasing deviation of the actual position and velocity of the robot from the prescribed values, caused by the deviation of the actual parameter values $p(\omega)$ from the chosen nominal values $p^{(0)}$, must be compensated by online control corrections $\Delta u(t)$, $t > t_0$. This requires usually extensive online state observations (measurements) and feedback control actions.

In order to determine a more reliable reference path $q = q(t)$, $t_0 \le t \le t_f$, in configuration space, being robust with respect to stochastic parameter variations, the a priori information (e.g., certain moments or parameters of the probability distribution of $p(\cdot)$) about the random variations of the vector $p(\omega)$ of model parameters of the robot and its working environment is taken into account already at the planning phase. Thus, instead of solving a deterministic trajectory planning problem with a fixed nominal parameter vector $p^{(0)}$, here, an optimal velocity profile $\beta^{(0)}$, $s_0 \le$

$s \le s_f$, and—in case of point-to-point control problems—also an optimal geometric path $q_e^{(0)}(s)$, $s_0 \le s \le s_f$, in configuration space is determined by using a stochastic optimization approach [25, 28–30, 36]. By means of $\beta^{(0)}(s)$ and $q_e^{(0)}(s)$, $s_0 \le s \le s_f$, we then find a more reliable, robust reference trajectory $q^{(0)}(t)$, $t_0 \le t \le t_f^{(0)}$, in configuration space. Applying now the so-called "inverse dynamics approach" [1, 4, 15], more reliable, robust open-loop controls $u^{(0)}(t)$, $t_0 \le t \le t_f^{(0)}$, are obtained. Moreover, by linearization of the dynamic equation of the robot in a neighborhood of $\left(u^{(0)}(t), q^{(0)}(t), E\left(p_M(\omega)|\mathcal{A}_{t_0}\right)\right)$, $t \ge t_0$, where $\mathcal{A}_{t_0}$ denotes the σ-algebra of informations up to the initial time point t_0, a control correction $\Delta u^{(0)}(t)$, $t \ge t_0$, is obtained which is related to the so-called feedback linearization of a system [4, 15, 37, 47].

At later moments (main correction time points) t_j,

$$t_0 < t_1 < t_2 < \cdots < t_{j-1} < t_j < \ldots, \tag{10.3}$$

further information on the parameters of the control system and its environment are available, e.g., by process observation, identification, calibration procedures, etc. Improvements $q^{(j)}(t), u^{(j)}(t), \Delta u^{(j)}(t)$, $t \ge t_j$, $j = 1, 2, \ldots$, of the preceding reference trajectory $q^{(j-1)}(t)$, open-loop control $u^{(j-1)}(t)$, and local control correction (feedback control) $\Delta u^{(j-1)}(t)$ can be determined by *replanning*, i.e., by optimal stochastic trajectory planning (OSTP) for the remaining time interval $t \ge t_j$, $j = 1, 2, \ldots$, and by using the information $\mathcal{A}_{t_j}$ on the robot and its working environment available up to the time point $t_j > t_0$, $j = 1, 2, \ldots$, see [16, 40, 41].

10.2 Optimal Trajectory Planning for Robots

According to [4, 34, 45], the dynamic equation for a robot is given by the following system of second-order differential equations

$$M\Big(p_D, q(t)\Big)\ddot{q}(t) + h\Big(p_D, q(t), \dot{q}(t)\Big) = u(t), \quad t \ge t_0, \tag{10.4a}$$

for the n-vector $q = q(t)$ of the robot or configuration coordinates $q_1, q_2, \ldots, q_n$. Here, $M = M(p_D, q)$ denotes the $n \times n$ inertia (or mass) matrix, and the vector function $h = h(p_D, q, \dot{q})$ is given by

$$h(p_D, q, \dot{q}) := C(p_D, q, \dot{q})\dot{q} + F_R(p_D, q, \dot{q}) + G(p_D, q), \tag{10.4b}$$

where $C(p_D, q, \dot{q}) = C(p_D, q)\dot{q}$, and $C(p_D, q) = \Big(C_{ijk}(p_D, q)\Big)_{1 \le i,j,k \le n}$ is the tensor of Coriolis and centrifugal terms, $F_R = F_R(p_D, q, \dot{q})$ denotes the vector of frictional forces and $G = G(p_D, q)$ is the vector of gravitational forces. Moreover,

$u = u(t)$ is the vector of controls, i.e., the vector of torques/forces in the joints of the robot. Standard representations of the friction term F_R are given [4, 19, 45] by

$$F_R(p_D, q, \dot{q}) := R_v(p_D, q)\dot{q}, \tag{10.4c}$$

$$F_R(p_D, q, \dot{q}) := R(p_D, q)sgn(\dot{q}), \tag{10.4d}$$

where $sgn(\dot{q}) := \Big(sgn(\dot{q}_1), \ldots, sgn(\dot{q}_n)\Big)^T$. In the first case (10.4c), $R_v = R_v(p_D, q)$ is the viscous friction matrix, and in the Coulomb approach (10.4d), $R = R(p_D, q) = \Big(R_i(p, q)\delta_{ij}\Big)$ is a diagonal matrix.

Remark 10.1 (*Inverse dynamics*) Reading the dynamic equation (10.4a) from the left to the right-hand side, hence, by inverse dynamics [1, 4, 15], the control function $u = u(t)$ may be described in terms of the trajectory $q = q(t)$ in configuration space.

The relationship between the so-called configuration space $\{q\}$ of robot coordinates $q = (q_1, \ldots, q_n)'$ and the work space $\{x\}$ of world coordinates (position and orientation of the end-effector) $x = (x_1, \ldots, x_n)'$ is represented by the kinematic equation

$$x = T(p_K, q). \tag{10.5}$$

As mentioned already in the introduction, p_D, p_K, denote the vectors of dynamic, kinematic parameters arising in the dynamic and kinematic equation (10.4a)–(10.4d), (10.5).

Remark 10.2 (*Linear parameterization of robots*) Note that the parameterization of a robot can be chosen, cf. [1, 4, 15], so that the dynamic and kinematic equation depend linearly on the parameter vectors p_D, p_K.

The objective of optimal trajectory planning is to determine [7, 8, 19, 35, 46] a control function $u = u(t)$, $t \geq t_0$, so that the cost functional

$$J\Big(u(\cdot)\Big) := \int_{t_0}^{t_f} L\Big(t, p_J, q(t), \dot{q}(t), u(t)\Big)\, dt + \phi\Big(t_f, p_J, q(t_f), \dot{q}(t_f)\Big) \tag{10.6}$$

is minimized, where the terminal time t_f may be given explicitly or implicitly, as, e.g., in minimum-time problems. Standard examples are, see, e.g., [34]:

(a) $\phi = 0$, $L = 1$ (minimum time),
(b) $\phi = 0$, $L =$ sum of potential, translatory, and rotational energy of the robot (minimum energy),
(c) $\phi = 0$, $L = \sum_{i=1}^{n} \Big(\dot{q}_i(t)u_i(t)\Big)^2$ (minimum fuel consumption),
(d) $\phi = 0$, $L = \sum_{i=1}^{n} \Big(u_i(t)\Big)^2$ (minimum force and moment).

Furthermore, an optimal control function $u^* = u^*(t)$ and the related optimal trajectory $q^* = q^*(t), t \geq t_0$, in configuration space must satisfy the dynamic equation (10.4a)–(10.4d) and the following constraints [7, 8, 10]:

(i) The initial conditions

$$q(t_0) = q_0(\omega), \quad \dot{q}(t_0) = \dot{q}_0(\omega) \tag{10.7}$$

Note that by means of the kinematic equation (10.5), the initial state $(q_0(\omega), \dot{q}_0(\omega))$ in configuration space can be represented by the initial state $\Big(x_0(\omega), \dot{x}_0(\omega)\Big)$ in work space.

(ii) The terminal conditions

$$\psi\Big(t_f, p, q(t_f), \dot{q}(t_f)\Big) = 0, \tag{10.8a}$$

e.g.,

$$q(t_f) = q_f(\omega), \quad \dot{q}(t_f) = \dot{q}_f(\omega). \tag{10.8b}$$

Again, by means of (10.5), $(q_f, \dot{q}_f)$ may be described in terms of the final state $(x_f, \dot{x}_f)$ in work space. Note that more general boundary conditions of this type may occur at some intermediate time points $t_0 < \tau_1 < \tau_2 < \cdots < \tau_r < t_f$.

(iii) Control constraints

$$u^{\min}(t, p) \leq u(t) \leq u^{\max}(t, p), \qquad t_0 \leq t \leq t_f \tag{10.9a}$$

$$g_I\Big(t, p, q(t), \dot{q}(t), u(t)\Big) \leq 0, \qquad t_0 \leq t \leq t_f \tag{10.9b}$$

$$g_{II}\Big(t, p, q(t), \dot{q}(t), u(t)\Big) = 0, \qquad t_0 \leq t \leq t_f. \tag{10.9c}$$

(iv) State constraints

$$S_I\Big(t, p, q(t), \dot{q}(t)\Big) \leq 0, \qquad t_0 \leq t \leq t_f \tag{10.10a}$$

$$S_{II}\Big(t, p, q(t), \dot{q}(t)\Big) = 0, \qquad t_0 \leq t \leq t_f. \tag{10.10b}$$

Using the kinematic equation (10.5), different types of obstacles in the work space can be described by (time-invariant) state constraints of the type (10.10a), (10.10b).

In robotics [35] often the following state constraints are used:

$$q_{\min}(p_C) \leq q(t) \leq q_{\max}(p_C), \qquad t_0 \leq t \leq t_f \tag{10.10c}$$

$$\dot{q}_{\min}(p_C) \leq \dot{q}(t) \leq \dot{q}_{\max}(p_C), \qquad t_0 \leq t \leq t_f, \tag{10.10d}$$

with certain vectors $q_{\min}, q_{\max}, \dot{q}_{\min}, \dot{q}_{\max}$ of (random) bounds.

A special constraint of the type (10.10b) occurs if the trajectory in work space

$$x(t) := T\Big(p_K, q(t)\Big) \tag{10.11}$$

should follow as precise as possible a geometric path in work space

$$x_e = x_e(p_x, s), \quad s_0 \le s \le s_f \tag{10.12}$$

being known up to a certain random parameter vector $p_x = p_x(\omega)$, which then is added to the total vector p of model parameters, cf. (10.4a), (10.4b).

Remark 10.3 In the following we suppose that the functions M, h, L, ϕ and T arising in (10.4a)–(10.4d), (10.5), (10.6) as well as the functions $\psi, g_I, g_{II}, S_I, S_{II}$ arising in the constraints (10.8a)–(10.10b) are sufficiently smooth.

10.3 Problem Transformation

Since the terminal time t_f may be given explicitly or implicitly, the trajectory $q(\cdot)$ in configuration space may have a varying domain $[t_0, t_f]$. Hence, in order to work with a given fixed domain of the unknown functions, the reference trajectory $q = q(t),\ t \ge t_0$, in configuration space is represented, cf. [19], by

$$q(t) := q_e\Big(s(t)\Big), \quad t \ge t_0. \tag{10.13a}$$

Here,

$$s = s(t), \quad t_0 \le t \le t_f \tag{10.13b}$$

is a strictly monotonous increasing transformation from the possibly varying time domain $[t_0, t_f]$ into a given fixed parameter interval $[s_0, s_f]$. For example, $s \in [s_0, s_f]$ may be the path parameter of a given path in work space, cf. (10.12). Moreover,

$$q_e = q_e(s), \quad s_0 \le s \le s_f \tag{10.13c}$$

denotes the so-called geometric path in configuration space.

Remark 10.4 In many more complicated industrial robot tasks such as grinding, welding, driving around difficult obstacles, complex assembly, etc., the geometric path $q_e(\cdot)$ in configuration space is predetermined offline [9, 16, 17] by a separate path planning procedure for $q_e = q_e(s),\ s_0 \le s \le s_f$, only. Hence, the trajectory planning/replanning is reduced then to the computation/adaptation of the transformation $s = s(t)$ along a given fixed path $q_e(\cdot) = q_e^{(0)}(\cdot)$.

Assuming that the transformation $s = s(t)$ is differentiable on $[t_0, t_f]$ with the exception of at most a finite number of points, we introduce now the so-called velocity

profile $\beta = \beta(s)$, $s_0 \le s \le s_f$, along the geometric path $q_e(\cdot)$ in configuration space by

$$\beta(s) := \dot{s}^2\Big(t(s)\Big) = \left(\frac{ds}{dt}\right)^2\Big(t(s)\Big), \tag{10.14}$$

where $t = t(s)$, $s_0 \le s \le s_f$, is the inverse of $s = s(t)$, $t_0 \le t \le t_f$. Thus, we have that

$$dt = \frac{1}{\sqrt{\beta(s)}}\, ds, \tag{10.15a}$$

and the time $t \ge t_0$ can be represented by the integral

$$t = t(s) := t_0 + \int_{s_0}^{s} \frac{d\sigma}{\sqrt{\beta(\sigma)}}. \tag{10.15b}$$

Using the integral transformation $\sigma := s_0 + (s - s_0)\rho, 0 \le \rho \le 1$, $t = t(s)$ may be also represented by

$$t(s) = t_0 + (s - s_0) \int_0^1 \frac{d\rho}{\sqrt{\beta\Big(s_0 + (s - s_0)\rho\Big)}}, \quad s \ge s_0. \tag{10.16a}$$

By numerical quadrature, i.e., by applying a certain numerical integration formula of order ν and having weights $a_0, a_1, a_2, \ldots, a_\nu$ to the integral in (10.16a), the time function $t = t(s)$ can be represented approximatively (with an $\varepsilon_0 > 0$) by

$$\tilde{t}(s) := t_0 + (s - s_0) \sum_{k=0}^{\nu} \frac{a_k}{\sqrt{\beta\left(s_0 + \varepsilon_0 + (s - s_0 - 2\varepsilon_0)\frac{k}{\nu}\right)}}, \quad s \ge s_0. \tag{10.16b}$$

In case of Simpson's rule ($\nu = 2$) we have that

$$\tilde{t}(s) := t_0 + \frac{s - s_0}{6}\left(\frac{1}{\sqrt{\beta(s_0 + \varepsilon_0)}} + \frac{4}{\sqrt{\beta\left(\frac{s + s_0}{2}\right)}} + \frac{1}{\sqrt{\beta(s - \varepsilon_0)}}\right). \tag{10.16c}$$

As long as the basic mechanical equations, the cost and constraint functions do **not** depend explicitly on time t, the transformation of the robot control problem from the time onto the s-parameter domain causes no difficulties. In the more general case one has to use the time representation (10.15b), (10.16a) or its approximates (10.16b), (10.16c).

Obviously, the terminal time t_f is given, cf. (10.15b), (10.16a), by

$$t_f = t(s_f) = t_0 + \int_{s_0}^{s_f} \frac{d\sigma}{\sqrt{\beta(\sigma)}} \tag{10.17}$$
$$= t_0 + (s_f - s_0) \int_0^1 \frac{d\rho}{\sqrt{\beta\Big(s_0 + (s_f - s_0)\rho\Big)}}.$$

10.3.1 Transformation of the Dynamic Equation

Because of (10.13a), (10.13b), we find

$$\dot{q}(t) = q_e'(s)\dot{s} \qquad \left(\dot{s} := \frac{ds}{dt}, q_e'(s) := \frac{dq_e}{ds}\right) \tag{10.18a}$$
$$\ddot{q}(t) = q_e'(s)\ddot{s} + q_e''(s)\dot{s}^2. \tag{10.18b}$$

Moreover, according to (10.14) we have that

$$\dot{s}^2 = \beta(s), \quad \dot{s} = \sqrt{\beta(s)}, \tag{10.18c}$$

and the differentiation of (10.18c) with respect to time t yields

$$\ddot{s} = \frac{1}{2}\beta'(s). \tag{10.18d}$$

Hence, (10.18a)–(10.18d) yields the following representation

$$\dot{q}(t) = q_e'(s)\sqrt{\beta(s)} \tag{10.19a}$$
$$\ddot{q}(t) = q_e'(s)\frac{1}{2}\beta'(s) + q_e''(s)\beta(s) \tag{10.19b}$$

of $\dot{q}(t), \ddot{q}(t)$ in terms of the new unknown functions $q_e(\cdot), \beta(\cdot)$.

Inserting now (10.19a), (10.19b) into the dynamic equation (10.4a), we find the equivalent relation

$$u_e\,(p_D, s; q_e(\cdot), \beta(\cdot)) = u(t) \;\text{ with }\; s = s(t), t = t(s), \tag{10.20a}$$

where the function u_e is defined by

$$u_e\Big(p_D, s; q_e(\cdot), \beta(\cdot)\Big) := M\Big(p_D, q_e(s)\Big)\left(\frac{1}{2}q_e'(s)\beta'(s) + q_e''(s)\beta(s)\right) + h\left(p_D, q_e(s), q_e'(s)\sqrt{\beta(s)}\right). \tag{10.20b}$$

The initial and terminal conditions (10.7)–(10.8b) are transformed, see (10.13a), (10.13b) and (10.19a), as follows

$$q_e(s_0) = \boldsymbol{q}_0(\omega), \quad q_e'(s_0)\sqrt{\beta(s_o)} = \dot{\boldsymbol{q}}_{\boldsymbol{0}}(\omega) \tag{10.21a}$$

$$\psi\Big(t(s_f), p, q_e(s_f), q_e'(s_f)\sqrt{\beta(s_f)}\Big) = 0 \tag{10.21b}$$

or

$$q_e(s_f) = \boldsymbol{q}_f(\omega), \quad q_e'(s_f)\sqrt{\beta(s_f)} = \dot{\boldsymbol{q}}_{\boldsymbol{f}}(\omega). \tag{10.21c}$$

Remark 10.5 In most cases we have the robot resting at time $t = t_0$ and $t = t_f$, i.e., $\dot{q}(t_0) = \dot{q}(t_f) = 0$, hence,

$$\beta(s_0) = \beta(s_f) = 0. \tag{10.21d}$$

10.3.2 Transformation of the Control Constraints

Using (10.13a), (10.13b), the control constraints (10.9a)–(10.9c) read in s-form as follows:

$$u^{\min}\Big(t(s), p_C\Big) \le u_e\Big(p_D, s; q_e(\cdot), \beta(\cdot)\Big) \le u^{\max}\Big(t(s), p_C\Big), \qquad s_0 \le s \le s_f \tag{10.22a}$$

$$g_I\left(t(s), p_C, q_e(s), q_e'(s)\sqrt{\beta(s)}, u_e\Big(p_D, s; q_e(\cdot), \beta(\cdot)\Big)\right) \le 0, \qquad s_0 \le s \le s_f \tag{10.22b}$$

$$g_{II}\left(t(s), p_C, q_e(s), q_e'(s)\sqrt{\beta(s)}, u_e\Big(p_D, s; q_e(\cdot), \beta(\cdot)\Big)\right) = 0, \quad s_0 \le s \le s_f, \tag{10.22c}$$

where $t = t(s) = t\Big(s; \beta(\cdot)\Big)$ or its approximation $t = \tilde{t}(s) = \tilde{t}\Big(s; \beta(\cdot)\Big)$ is defined by (10.15b), (10.16a)–(10.16c).

Remark 10.6

(I) In the important case that the bounds for $u = u(t)$ depend on the system state $\Big(q(t), \dot{q}(t)\Big)$ in configuration space,i.e.,

$$\begin{aligned} u^{\min}(t, p_C) &:= u^{\min}\Big(p_C, q(t), \dot{q}(t)\Big), \\ u^{\max}(t, p_C) &:= u^{\max}\Big(p_C, q(t), \dot{q}(t)\Big) \end{aligned} \tag{10.23a}$$

condition (10.22a) is reduced to

$$\begin{aligned} u^{\min}\Big(p_C, q_e(s), q_e'(s)\sqrt{\beta(s)}\Big) \le u_e\Big(p_D, s; q_e(\cdot), \beta(\cdot)\Big) \\ \le u^{\max}\Big(p_C, q_e(s), q_e'(s)\sqrt{\beta(s)}\Big), \quad s_0 \le s \le s_f. \end{aligned} \tag{10.23b}$$

(II) If the bounds for $u(t)$ in (10.23a) do not depend on the velocity $\dot{q}(t)$ in configuration space, and the geometric path $q_e(s) = q_e(s)$, $s_0 \le s \le s_f$, in configuration space is known in advance, then the bounds

$$\begin{aligned} u^{\min}\Big(p_C, q_e(s)\Big) &= \tilde{u}^{\min}(p_C, s) \\ u^{\max}\Big(p_C, q_e(s)\Big) &= \tilde{u}^{\max}(p_C, s), \quad s_0 \le s \le s_f \end{aligned} \tag{10.23c}$$

depend on (p_C, s) only.
Bounds of the type (10.23c) for the control function $u(t)$ may be taken into account as an approximation of the more general bounds in (10.22a).

10.3.3 Transformation of the State Constraints

Applying the transformations (10.13a), (10.13b), (10.18a) and (10.15b) to the state constraints (10.10a), (10.10b), we find the following s-form of the state constraints:

$$S_I\Big(t(s), p_C, q_e(s), q_e'(s)\sqrt{\beta(s)}\Big) \le 0, \qquad s_0 \le s \le s_f \tag{10.24a}$$

$$S_{II}\Big(t(s), p_C, q_e(s), q_e'(s)\sqrt{\beta(s)}\Big) = 0, \qquad s_0 \le s \le s_f. \tag{10.24b}$$

Obviously, the s-form of the special state constraints (10.10c), (10.10d) read

$$q^{\min}(p_C) \le q_e(s) \le q^{\max}(p_C), \qquad s_0 \le s \le s_f, \tag{10.24c}$$

$$\dot{q}^{\min}(p_C) \le q_e'(s)\sqrt{\beta(s)} \le \dot{q}^{\max}(p_C), \qquad s_0 \le s \le s_f. \tag{10.24d}$$

In the case that the end-effector of the robot has to follow a given path (10.12) in work space, Eq. (10.24b) reads

$$T\Big(p_K, q_e(s)\Big) - x_e(p_x, s) = 0, \quad s_0 \le s \le s_f, \tag{10.24e}$$

with the parameter vector p_x describing possible uncertainties in the selection of the path to be followed by the roboter in work space.

10.3.4 Transformation of the Objective Function

Applying the integral transformation $t = t(s), dt = \dfrac{ds}{\sqrt{\beta(s)}}$ to the integral in the representation (10.6) of the objective function $J = J\Big(u(\cdot)\Big)$, and transforming also the terminal costs, we find the following s-form of the objective function:

$$J\Big(u(\cdot)\Big) = \int_{s_0}^{s_f} L\left(t(s),\ p_J, q_e(s), q_e'(s)\sqrt{\beta(s)}, u_e\Big(p_D, s; q_e(\cdot), \beta(\cdot)\Big)\right) \frac{ds}{\sqrt{\beta(s)}}$$
$$+ \phi\left(t(s_f), p_J, q_e(s_f), q_e'(s_f)\sqrt{\beta(s_f)}\right). \tag{10.25a}$$

Note that $\beta(s_f) = 0$ holds in many practical situations.

For the class of time-minimum problems we have that

$$J\Big(u(\cdot)\Big) := t_f - t_0 = \int_{t_0}^{t_f} dt = \int_{s_0}^{s_f} \frac{ds}{\sqrt{\beta(s)}}. \tag{10.25b}$$

Optimal deterministic trajectory planning (OSTP). By means of the $t - s$-transformation onto the fixed s–parameter domain $[s_0, s_f]$, the optimal control problem (10.4a)–(10.4d), (10.6)–(10.12) is transformed into a variational problem for finding, see (10.13a)–(10.13c) and (10.14), an optimal velocity profile $\beta(s)$ and an optimal geometric path $q_e(s)$, $s_0 \le s \le s_f$. In the deterministic case, i.e., if the parameter vector p is assumed to be known, then for the numerical solution of the resulting *optimal deterministic trajectory planning problem* several efficient solution techniques are available, cf. [7, 8, 10, 19, 25, 30, 46].

10.4 OSTP—Optimal Stochastic Trajectory Planning

In the following we suppose that the initial and terminal conditions (10.21d) hold, i.e.,

$$\beta_0 = \beta(s_0) = \beta_f = \beta(s_f) = 0 \text{ or } \dot{q}(t_0) = \dot{q}(t_f) = 0.$$

Based on the $(t-s)$–transformation described in Sect. 10.3, and relying on the inverse dynamics approach, the *robot control problem* (10.6), (10.7)–(10.8b), (10.9a)–(10.9c), (10.10a)–(10.10c) can be represented now by a *variational problem* for $\Big(q_e(\cdot), \beta(\cdot)\Big), \beta(\cdot)$, resp., given in the following. Having $\Big(q_e(\cdot), \beta(\cdot)\Big), \beta(\cdot)$, resp., a reference trajectory and a feedforward control can then be constructed.

(A) *Time-invariant case* (autonomous systems)
If the objective function and the constraint functions do not depend explicitly on time t, then the optimal control problem takes the following equivalent s-forms:

$$\min \int_{s_0}^{s_f} L^J\Big(p_J, q_e(s), q_e'(s), q_e''(s), \beta(s), \beta'(s)\Big)\, ds + \phi^J\Big(p_J, q_e(s_f)\Big) \tag{10.26a}$$

s.t.

$$f_I^u\Big(p, q_e(s), q_e'(s), q_e''(s), \beta(s), \beta'(s)\Big) \le 0, \qquad s_0 \le s \le s_f \tag{10.26b}$$

$$f_{II}^u\Big(p, q_e(s), q_e'(s), q_e''(s), \beta(s), \beta'(s)\Big) = 0, \qquad s_0 \le s \le s_f \tag{10.26c}$$

$$f_I^S\Big(p, q_e(s), q_e'(s), \beta(s)\Big) \le 0, \qquad s_0 \le s \le s_f \tag{10.26d}$$

$$f_{II}^S\Big(p, q_e(s), q_e'(s), \beta(s)\Big) = 0, \qquad s_0 \le s \le s_f \tag{10.26e}$$

$$\beta(s) \ge 0, \qquad s_0 \le s \le s_f \tag{10.26f}$$

$$q_e(s_0) = \mathbf{q}_0(\omega), \quad q_e'(s_0)\sqrt{\beta(s_0)} = \dot{\boldsymbol{q}}_{\mathbf{0}}(\omega) \tag{10.26g}$$

$$q_e(s_f) = \mathbf{q}_f(\omega), \quad \beta(s_f) = \beta_f. \tag{10.26h}$$

Under condition (10.21d), a more general version of the terminal condition (10.26h) reads, cf. (10.21b),

$$\psi\Big(p, q_e(s_f)\Big) = 0, \quad \beta(s_f) = \beta_f := 0. \tag{10.26h'}$$

Here,

$$L^J = L^J\Big(p_J, q_e, q_e', q_e'', \beta, \beta'\Big), \quad \phi^J = \phi^J(p_J, q_e) \tag{10.27a}$$

$$f_I^u = f_I^u(p, q_e, q_e', q_e'', \beta, \beta'), \quad f_{II}^u = f_{II}^u(p, q_e, q_e', q_e'', \beta, \beta') \tag{10.27b}$$

$$f_I^S = f_I^S(p, q_e, q_e', \beta), \quad f_{II}^S = f_{II}^S(p, q_e, q_e', \beta) \tag{10.27c}$$

are the functions representing the s-form of the objective function (10.25a), the constraint functions in the control constraints (10.22a)–(10.22c), and in the state constraints (10.24a)–(10.24e), respectively. Define then f^u and f^S by

$$f^u := \begin{pmatrix} f_I^u \\ f_{II}^u \end{pmatrix}, \quad f^S := \begin{pmatrix} f_I^S \\ f_{II}^S \end{pmatrix}. \tag{10.27d}$$

(B) *Time-varying case* (non autonomous systems)
If the time t occurs explicitly in the objective and/or in some of the constraints of the robot control problem, then, using (10.15a), (10.15b), (10.16a)–(10.16c), we have that $t = t\Big(s; t_0, s_0, \beta(\cdot)\Big)$, and the functions (10.27a)–(10.27d) and ψ may depend then also on $\Big(s, t_0, s_0, \beta(\cdot)\Big)$, $\Big(s_f, t_0, s_0, \beta(\cdot)\Big)$, resp., hence,

$$\begin{aligned}
L^J &= L^J\Big(s, t_0, s_0, \beta(\cdot), p_J, q_e, q_e', q_e'', \beta, \beta'\Big) \\
\phi^J &= \phi^J\Big(s_f, t_0, s_0, \beta(\cdot), p_J, q_e\Big) \\
f^u &= f^u\Big(s, t_0, s_0, \beta(\cdot), p, q_e, q_e', q_e'', \beta, \beta'\Big) \\
f^S &= f^S\Big(s, t_0, s_0, \beta(\cdot), p, q_e, q_e', \beta\Big) \\
\psi &= \psi\Big(s_f, t_0, s_0, \beta(\cdot), p, q_e\Big).
\end{aligned}$$

In order to get a reliable optimal geometric path $q_e^* = q_e^*(s)$ in configuration space and a reliable optimal velocity profile $\beta^* = \beta^*(s)$, $s_0 \le s \le s_f$, being robust with respect to random parameter variations of $p = p(\omega)$, the variational problem (10.26a)–(10.26h) under stochastic uncertainty must be replaced by an appropriate *deterministic substitute problem* which is defined according to the following principles [21–24, 30], cf. also [20, 21, 24, 26, 27].

Assume first that the a priori information about the robot and its environment up to time t_0 is described by means of a σ–algebra $\mathcal{A}_{t_0}$, and let then

$$P_{p(\cdot)}^{(0)} = P_{p(\cdot)}\Big(\cdot|\mathcal{A}_{t_0}\Big) \tag{10.28}$$

denote the a priori distribution of the random vector $p = p(\omega)$ given $\mathcal{A}_{t_0}$.

Depending on the decision theoretical point of view, different approaches are possible, e.g., reliability-based substitute problems, belonging essentially to one of the following two basic classes of substitute problems:

(I) Risk(recourse)-constrained minimum expected cost problems
(II) Expected total cost-minimum problems.

Substitute problems are constructed by selecting certain scalar or vectorial loss or cost functions

$$\gamma_I^u, \gamma_{II}^u, \gamma_I^S, \gamma_{II}^S, \gamma^\psi, \ldots \tag{10.29a}$$

evaluating the violation of the random constraints (10.26b), (10.26c), (10.26d), (10.26e), (10.26h'), respectively.

In the following all expectations are conditional expectations with respect to the a priori distribution $P_{p(\cdot)}^{(0)}$ of the random parameter vector $p(\omega)$. Moreover, the following compositions are introduced:

$$f_\gamma^u := \begin{pmatrix} \gamma_I^u \circ f_I^u \\ \gamma_{II}^u \circ f_{II}^u \end{pmatrix}, \quad f_\gamma^S := \begin{pmatrix} \gamma_I^S \circ f_I^S \\ \gamma_{II}^S \circ f_{II}^S \end{pmatrix} \tag{10.29b}$$

$$\psi_\gamma := \gamma^\psi \circ \psi. \tag{10.29c}$$

Now the two basic types of substitute problems are described.

(I) *Risk(recourse)-based minimum expected cost problems*
Minimizing the expected (primal) costs $E\left(J(u(\cdot))|\mathcal{A}_{t_0}\right)$, and demanding that the risk, i.e., the expected (recourse) costs arising from the violation of the constraints of the variational problem (10.26a)–(10.26h) do not exceed given upper bounds, in the **time-invariant case** we find the following substitute problem:

$$\min \int_{s_0}^{s_f} E\left(L^J\left(p_J, q_e(s), q_e'(s), q_e''(s), \beta(s), \beta'(s)\right)|\mathcal{A}_{t_0}\right) ds \tag{10.30a}$$
$$+ E\left(\phi^J\left(p_J, q_e(s_f)\right)|\mathcal{A}_{t_0}\right)$$

s.t.

$$E\left(f_\gamma^u\left(p, q_e(s), q_e'(s), q_e''(s), \beta(s), \beta'(s)\right)|\mathcal{A}_{t_0}\right) \le \Gamma^u, \ s_0 \le s \le s_f \tag{10.30b}$$

$$E\left(f_\gamma^S\left(p, q_e(s), q_e'(s), \beta(s)\right)|\mathcal{A}_{t_0}\right) \le \Gamma^S, \quad s_0 \le s \le s_f \tag{10.30c}$$

$$\beta(s) \ge 0, \quad s_0 \le s \le s_f \tag{10.30d}$$

$$q_e(s_0) = \overline{q}_0, \quad q_e'(s_0)\sqrt{\beta(s_0)} = \dot{\overline{q}}_0 \tag{10.30e}$$

$$q_e(s_f) = \overline{q}_f \ (\text{if } \phi^J = 0), \quad \beta(s_f) = \beta_f, \tag{10.30f}$$

and the more general terminal condition (10.26h') is replaced by

$$\beta(s_f) = \beta_f := 0, \quad E\left(\psi_\gamma\left(p, q_e(s_f)\right)|\mathcal{A}_{t_0}\right) \leq \Gamma_\psi. \tag{10.30f'}$$

Here,

$$\Gamma^u = \Gamma^u(s), \quad \Gamma^S = \Gamma^S(s), \quad \Gamma_\psi = \Gamma_\psi(s) \tag{10.30g}$$

denote scalar or vectorial upper risk bounds which may depend on the path parameter $s \in [s_0, s_f]$. Furthermore, the initial, terminal values $\overline{q}_0, \overline{\dot{q}}_0, \overline{q}_f$ in (10.30e), (10.30f) are determined according to one of the following relations:

a.

$$\overline{q}_0 := \hat{q}(t_0), \quad \overline{\dot{q}}_0 := \hat{\dot{q}}(t_0), \quad \overline{q}_f := \hat{q}(t_f), \tag{10.30h}$$

where $\left(\hat{q}(t), \hat{\dot{q}}(t)\right)$ denotes an estimate, observation, etc., of the state in configuration space at time t;

b.

$$\begin{aligned} \overline{q}_0 &:= E\left(q_0(\omega)|\mathcal{A}_{t_0}\right), \quad \overline{\dot{q}}_0 := E\left(\dot{q}_0(\omega)|\mathcal{A}_{t_0}\right), \\ \overline{q}_f = \overline{q_f}^{(0)} &:= E\left(q_f(\omega)|\mathcal{A}_{t_0}\right), \end{aligned} \tag{10.30i}$$

where $q_0(\omega), \dot{q}_0(\omega)$ is a random initial position, and $q_f(\omega)$ is a random terminal position.

Having corresponding information about initial and terminal values $x_0, \dot{x}_0, x_f$ in work space, related equations for $q_0, \dot{q}_0, q_f$ may be obtained by means of the kinematic equation (10.5).

Remark 10.7 (*Average constraints*) Taking the average of the pointwise constraints (10.30b), (10.30c) with respect to the path parameter $s, s_0 \leq s \leq s_f$, we get the simplified integrated constraints

$$\int_{s_0}^{s_f} E\left(f_\gamma^u\left(p, q_e, (s), q_e'(s), q_e''(s), \beta(s), \beta'(s)\right)|\mathcal{A}_{t_0}\right) ds \leq \tilde{\Gamma}^u \tag{10.30b'}$$

$$\int_{s_0}^{s_f} E\left(f_\gamma^S\left(p, q_e(s), q_e'(s), \beta(s)\right)|\mathcal{A}_{t_0}\right) ds \leq \tilde{\Gamma}^S. \tag{10.30c'}$$

Remark 10.8 (*Generalized area of admissible motion*) In generalization of the admissible area of motion [19, 25, 33] for path planning problems with a prescribed geometrical path $q_e(\cdot) = \overline{q}_e(\cdot)$ in configuration space, for point-to-point problems the constraints (10.30b)–(10.30i) define for each path point $s, s_0 \leq s \leq s_f$, a generalized admissible area of motion for the vector

$$\chi(s) := \Big(q_e(s), q_e'(s), q_e''(s), \beta(s), \beta'(s)\Big), \quad s_0 \le s \le s_f, \tag{10.30j}$$

including information about the magnitude $\Big(\beta(s), \beta'(s)\Big)$ of the motion as well as information about the direction $\Big(q_e(s), q_e'(s), q_e''(s)\Big)$ of the motion.

Remark 10.9 (*Problems with Chance Constraints*) Substitute problems having chance constraints are obtained if the loss functions γ^u, γ^S for evaluating the violation of the inequality constraints in (10.26a)–(10.26h), (10.26h') are 0–1 functions, cf. [25].

To give a characteristic example, we demand that the control, state constraints (10.22a), (10.24c), (10.24d), resp., have to be fulfilled at least with probability $\alpha_u, \alpha_q, \alpha_{\dot q}$ for $s_0 \le s \le s_f$, hence,

$$P\Big(u^{\min}(p_C) \le u_e\Big(p_D, s; q_e(\cdot), \beta(\cdot)\Big) \le u^{\max}(p_C)|\mathcal{A}_{t_0}\Big) \ge \alpha_u, \tag{10.31a}$$

$$P\Big(q^{\min}(p_C) \le q_e(s) \le q^{\max}(p_C)|\mathcal{A}_{t_0}\Big) \ge \alpha_q, \tag{10.31b}$$

$$P\Big(\dot q^{\min}(p_C) \le q_e'(s)\sqrt{\beta(s)} \le \dot q^{\max}(p_C)|\mathcal{A}_{t_0}\Big) \ge \alpha_{\dot q}. \tag{10.31c}$$

Sufficient conditions for the chance constraints (10.31a)–(10.31c) can be obtained by applying certain probability inequalities, see [25]. Defining

$$\begin{aligned} u^c(p_C) &:= \frac{u^{\max}(p_C) + u^{\min}(p_C)}{2}, \\ \rho_u(p_C) &:= \frac{u^{\max}(p_C) - u^{\min}(p_C)}{2}, \end{aligned} \tag{10.31d}$$

then a sufficient conditions for (10.31a) reads, cf. [25],

$$\begin{aligned} &E\Big(tr B\rho_u(p_C)_d^{-1}\Big(u_e - u^c(p_C)\Big)\Big(u_e - u^c(p_C)\Big)^T\rho_u(p_C)_d^{-1}|\mathcal{A}_{t_0}\Big) \\ &\quad \le 1 - \alpha_u, \quad s_0 \le s \le s_f, \end{aligned} \tag{10.31e}$$

where $u_e = u_e\Big(p_D, s; q_e(\cdot), \beta(\cdot)\Big)$ and $\rho_u(p_C)_d$ denotes the diagonal matrix containing the elements of $\rho_u(p_C)$ on its diagonal. Moreover, B denotes a positive definite matrix such that $z^T B z \ge 1$ for all vectors z such that $\|z\|_\infty \ge 1$. Taking, e.g., $B = I$, (10.31e) reads

$$\begin{aligned} &E\Big(\|\rho_u(p_C)_d^{-1}\Big(u_e\Big(p_D, s; q_e(\cdot), \beta(\cdot)\Big) - u^c(p_C)\Big)\|^2|\mathcal{A}_{t_0}\Big) \\ &\quad \le 1 - \alpha_u, \quad s_0 \le s \le s_f. \end{aligned} \tag{10.31f}$$

Obviously, similar sufficient conditions may be derived for (10.31b), (10.31c).

We observe that the above class of risk-based minimum expected cost problems for the computation of $\Big(q_e(\cdot), \beta(\cdot)\Big), \beta(\cdot)$, resp., is represented completely by the following set of

$$initial\ parameters\ \zeta_0 :\ t_0, s_0, \overline{q}_0, \dot{\overline{q}}_0, P^{(0)}_{p(\cdot)} \text{ or } \nu_0 \tag{10.32a}$$

and

$$terminal\ parameters\ \zeta_f :\ t_f, s_f, \beta_f, \overline{q}_f. \tag{10.32b}$$

In case of problems with a given geometric path $q_e = q_e(s)$ in configuration space, the values q_0, q_f may be deleted. Moreover, approximating the expectations in (10.30a)–(10.30f), (10.30f') by means of Taylor expansions with respect to the parameter vector p at the conditional mean

$$\overline{p}^{(0)} := E\Big(p(\omega)|\mathcal{A}_{t_0}\Big), \tag{10.32c}$$

the a priori distribution $P^{(0)}_{p(\cdot)}$ may be replaced by a certain vector

$$\nu_0 := \left(E\Big(\prod_{k=1}^{r} p_l(\omega)|\mathcal{A}_{t_0}\Big)_{(l_1,\ldots,l_r)\in\Lambda}\right) \tag{10.32d}$$

of a priori moments of $p(\omega)$ with respect to $\mathcal{A}_{t_0}$.

Here, Λ denotes a certain finite set of multiple indices $(l_1, \ldots, l_\tau), r \leq 1$.

Of course, in the *time-variant case* the functions $L^J, \phi^J, f^u, f^S, \psi$ as described in item *(B)* have to be used. Thus, t_0, t_f occur then explicitly in the parameter list (10.32a), (10.32b).

(II) *Expected total cost-minimum problem*

Here, the total costs arising from violations of the constraints in the variational problem (10.30a)–(10.30f), (10.30f') are added to the (primary) costs arising along the trajectory, to the terminal costs, respectively. Of course, corresponding weight factors may be included in the cost functions (10.29a). Taking expectations with respect to $\mathcal{A}_{t_0}$, in the *time-invariant case* the following substitute problem is obtained:

$$\min \int_{s_0}^{s_f} E\left(L^J_\gamma\Big(p, q_e(s), q'_e(s), q''_e(s), \beta(s), \beta'(s)\Big)|\mathcal{A}_{t_0}\right) ds$$
$$+E\left(\phi^J_\gamma\Big(p, q_e(s_f)\Big)|\mathcal{A}_{t_0}\right) \tag{10.33a}$$

s.t.

$$\beta(s) \geq 0, \quad s_0 \leq s \leq s_f \tag{10.33b}$$

$$q_e(s_0) = \overline{q}_0, \quad q_e'(s_0)\sqrt{\beta(s_0)} = \overline{\dot{q}}_0 \tag{10.33c}$$

$$q_e(s_f) = \overline{q}_f \quad (\text{if } \phi_\gamma^J = 0), \beta(s_f) = \beta_f, \tag{10.33d}$$

where $L_\gamma^J, \phi_\gamma^J$ are defined by

$$L_\gamma^J := L^J + \nu^{uT} f_\gamma^u + \nu^{ST} f_\gamma^s \tag{10.33e}$$

$$\phi_\gamma^J := \phi^J \quad or \quad \phi_\gamma^J := \phi^J + \nu_\psi^T \psi_\gamma, \tag{10.33f}$$

and $\nu_I^u, \nu_{II}^u, \nu_I^S, \nu_{II}^S, \nu_\psi \geq 0$ are certain nonnegative (vectorial) scale factors which may depend on the path parameter s.

We observe that also in this case the initial/terminal parameters characterizing the second class of substitute problems (10.33a)–(10.33f) are given again by (10.32a), (10.32b).

In the *time-varying case* the present substitute problems of *class II* reads

$$\min \int_{s_0}^{s_f} E\left(L_\gamma^J\Big(s, t_0, s_0, \beta(\cdot), p_J, q_e(s), q_e'(s, q_e''(s), \beta(s), \beta'(s)\Big)|\mathcal{A}_{t_0}\right) ds$$
$$+ E\left(\phi_\gamma^J\Big(s_f, t_0, s_0, \beta(\cdot), p_J, q_e(s_f)\Big)|\mathcal{A}_{t_0}\right) \tag{10.34a}$$

s.t.

$$\beta(s) \geq 0, \quad s_0 \leq s \leq s_f \tag{10.34b}$$

$$q_e(s_0) = q_0, \quad q_e'(s_0)\sqrt{\beta(s_0)} = \dot{q}_0 \tag{10.34c}$$

$$q_e(s_f) = q_f \quad (\text{ if } \phi_\gamma^J = 0), \beta(s_f) = \beta_f. \tag{10.34d}$$

Remark 10.10 (*Mixtures of (I), (II)*) Several mixtures of the classes (I) and (II) of substitute problems are possible.

10.4.1 Computational Aspects

The following techniques are available for solving substitutes problems of type (I), (II):

(a) *Reduction to a finite-dimensional parameter optimization problem*
Here, the unknown functions $\Big(q_e(\cdot), \beta(\cdot)\Big)$ or $\beta(\cdot)$ are approximated by a linear combination

$$q_e(s) := \sum_{l=1}^{l_q} \hat{q}_l B_l^q(s), \quad s_0 \le s \le s_f \tag{10.35a}$$

$$\beta(s) := \sum_{l=1}^{l_\beta} \hat{\beta}_l B_l^\beta(s), \quad s_0 \le s \le s_f, \tag{10.35b}$$

where $B_l^q = B_l^q(s)$, $B_l^\beta = B_l^\beta(s)$, $s_0 \le s \le s_f$, $l = 1, \ldots, l_q(l_\beta)$, are given basis functions, e.g., B-splines, and $\hat{q}_l$, $\hat{\beta}_l$, $l = 1, \ldots, l_q(l_\beta)$, are vectorial, scalar coefficients. Putting (10.35a), (10.35b) into (10.30a)–(10.30f), (10.30f'), (10.33a)–(10.33f), resp., a semiinfinite optimization problem is obtained. If the inequalities involving explicitly the path parameter s, $s_0 \le s \le s_f$, are required for a finite number N of parameter values $s_1, s_2, \ldots, s_N$ only, then this problem is reduced finally to a finite-dimensional parameter optimization problem which can be solved now numerically by standard mathematical programming routines or search techniques. Of course, a major problem is the approximative computation of the conditional expectations which is done essentially by means of Taylor expansion with respect to the parameter vector p at $\overline{p}^{(0)}$. Consequently, several conditional moments have to be determined (online, for stage $j \ge 1$). For details, see [28–30, 36] and the program package "OSTP" [3].

(b) *Variational techniques*
Using methods from calculus of variations, necessary and—in some cases—also sufficient conditions in terms of certain differential equations may be derived for the optimal solutions $\left(q_e^{(0)}, \beta^{(0)}\right)$, $\beta^{(0)}$, resp., of the variational problems (10.30a)–(10.30f), (10.30f'), (10.33a)–(10.33f). For more details, see [36].

(c) *Linearization methods*
Here, we assume that we already have an approximative optimal solution $\left(\overline{q}_e(s), \overline{\beta}(s)\right)$, $s_0 \le s \le s_f$, of the substitute problem (10.30a)–(10.30f), (10.30f') or (10.33a)–(10.33f) under consideration. For example, an approximative optimal solution $\left(\overline{q}_e(\cdot), \overline{\beta}(\cdot)\right)$ can be obtained by starting from the deterministic substitute problem obtained by replacing the random parameter vector $p(\omega)$ just by its conditional mean $\overline{p}^{(0)} := E\left(p(\omega)|\mathcal{A}_{t_0}\right)$.
Given an approximate optimal solution $\left(\overline{q}_e(\cdot), \overline{\beta}(\cdot)\right)$ of substitute problem (I) or (II), the unknown optimal solution $\left(q_e^{(0)}(\cdot), \beta^{(0)}(\cdot)\right)$ to be determined is represented by

$$q_e^{(0)}(s) := \overline{q}_e(s) + \Delta q_e(s), \ s_0 \le s \le s_f \tag{10.36a}$$

$$\beta^{(0)}(s) := \overline{\beta}(s) + \Delta\beta(s), s_0 \le s \le s_f, \tag{10.36b}$$

where $\Big(\Delta q_e(s), \Delta\beta(s)\Big), s_0 \le s \le s_f$, are certain (small) correction terms. In the following we assume that the changes $\Delta q_e(s), \Delta\beta(s)$ and their first and resp. second order derivatives $\Delta q_e'(s), \Delta q_e''(s), \Delta\beta'(s)$ are small.
We observe that the function arising in the constraints and in the objective of (10.30a)–(10.30f), (10.30f'), (10.33a)–(10.33f), resp., are of the following type:

$$\begin{aligned}
&\overline{g}^{(0)}\Big(q_e(s), q_e'(s), q_e''(s), \beta(s), \beta'(s)\Big) \\
&:= E\left(g\Big(p(\omega), q_e(s), q_e'(s),\ q_e''(s),\ \beta(s), \beta'(s)\Big)|\mathcal{A}_{t_0}\right), \qquad (10.37a)\\
&\overline{\phi}^{(0)}\Big(q_e(s_f)\Big) := E\left(\phi\Big(p(\omega), q_e(s_f)\Big)|\mathcal{A}_{t_0}\right) \qquad (10.37b)
\end{aligned}$$

and

$$\overline{F}^{(0)}\Big(q_e(\cdot), \beta(\cdot)\Big) := \int_{s_0}^{s_f} \overline{g}^{(0)}\Big(q_e(s), q_e'(s), q_e''(s), \beta(s), \beta'(s)\Big)\, ds \qquad (10.37c)$$

with certain functions g, ϕ. Moreover, if for simplification the pointwise (cost-) constraints (10.30b), (10.30c) are averaged with respect to the path parameter $s, s_0 \le s_f$, then also constraint functions of the type (10.37c) arise, see (10.30b'), (10.30c').
By means of first-order Taylor expansion of g, ϕ with respect to $(\Delta q_e(s), \Delta q_e'(s), \Delta q_e''(s), \Delta\beta(s), \Delta\beta'(s))$ at $\Big(\overline{q}_e(s), \overline{q}_e'(s), \overline{q}_e''(s), \overline{\beta}(s), \overline{\beta}'(s)\Big), s_0 \le s \le s_f$, we find then the following approximations

$$\begin{aligned}
&\overline{g}^{(0)}\Big(q_e(s), q_e'(s), q_e''(s), \beta(s), \beta'(s)\Big) \approx \overline{g}^{(0)}\Big(\overline{q}_e(s), \overline{q}_e'(s), \overline{q}_e''(s), \overline{\beta}(s), \overline{\beta}'(s)\Big) \\
&+ \overline{A}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s)^T \Delta q_e(s) + \overline{B}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s)^T \Delta q_e'(s) + \overline{C}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s)^T \Delta q_e''(s) \\
&+ \overline{D}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s)\Delta\beta(s) + \overline{E}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s)\Delta\beta'(s) \qquad (10.38a)
\end{aligned}$$

and

$$\overline{\phi}^{(0)}\Big(q_e(s_f)\Big) \approx \overline{\phi}^{(0)}\Big(\overline{q}_e(s_f)\Big) + \overline{a}^{(0)}_{\phi,\overline{q}_e}(s_f)^T \Delta q_e(s_f), \qquad (10.38b)$$

where the expected sensitivities of g, ϕ with respect to q, q', q'', β and β' are given by

$$\overline{A}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s) := E\left(\nabla_q g\Big(p(\omega),\overline{q}_e(s),\overline{q}'_e(s),\overline{q}''_e(s),\overline{\beta}(s),\overline{\beta}'(s)\Big)|\mathcal{A}_{t_0}\right) \tag{10.38c}$$

$$\overline{B}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s) := E\left(\nabla_{q'} g\Big(p(\omega),\overline{q}_e(s),\overline{q}'_e(s),\overline{q}''_e(s),\overline{\beta}(s),\overline{\beta}'(s)\Big)|\mathcal{A}_{t_0}\right) \tag{10.38d}$$

$$\overline{C}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s) := E\left(\nabla_{q''} g\Big(p(\omega),\overline{q}_e(s),\overline{q}'_e(s),\overline{q}''_e(s),\overline{\beta}(s),\overline{\beta}'(s)\Big)|\mathcal{A}_{t_0}\right) \tag{10.38e}$$

$$\overline{D}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s) := E\left(\frac{\partial g}{\partial \beta}\Big(p(\omega),\overline{q}_e(s),\overline{q}'_e(s),\overline{q}''_e(s),\overline{\beta}(s),\overline{\beta}'(s)\Big)|\mathcal{A}_{t_0}\right) \tag{10.38f}$$

$$\overline{E}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s) := E\left(\frac{\partial g}{\partial \beta'}\Big(p(\omega),\overline{q}_e(s),\overline{q}'_e(s),\overline{q}''_e(s),\overline{\beta}(s),\overline{\beta}'(s)\Big)|\mathcal{A}_{t_0}\right) \tag{10.38g}$$

$$\overline{a}^{(0)}_{\phi,\overline{q}_e}(s_f) := E\left(\nabla_q \phi\Big(p(\omega),\overline{q}_e(s)\Big)|\mathcal{A}_{t_0}\right), s_0 \le s \le s_f. \tag{10.38h}$$

As mentioned before, cf. (10.32c), (10.32d), the expected values $\overline{g}^{(0)}, \overline{\phi}^{(0)}$ in (10.38a), (10.38b) and the expected sensitivities defined by (10.38c)–(10.38h) can be computed approximatively by means of Taylor expansion with respect to p at $\overline{p}^{(0)} = E\Big(p(\omega)|\mathcal{A}_{t_0}\Big)$.

According to (10.38a), and using partial integration, for the total costs $\overline{F}^{(0)}$ along the path we get the following approximation:

$$\begin{aligned}
\overline{F}^{(0)} &\approx \int_{s_0}^{s_f} \overline{g}^{(0)}\Big(\overline{q}_e(s),\overline{q}'_e(s),\overline{q}''_e(s),\overline{\beta}(s),\overline{\beta}'(s)\Big)\,ds \qquad (10.39a)\\
&+\int_{s_0}^{s_f} \overline{G}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s)^T \Delta q_e(s)\,ds + \int_{s_0}^{s_f} \overline{H}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s)\Delta\beta(s)\,ds\\
&+\Big(\overline{B}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s_f) - \overline{C}^{(0)'}_{g,\overline{q}_e,\overline{\beta}}(s_f)\Big)^T \Delta q_e(s_f)\\
&+\Big(-\overline{B}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s_0) + \overline{C}^{(0)'}_{g,\overline{q}_e,\overline{\beta}}(s_0)\Big)^T \Delta q_e(s_0)\\
&+\overline{C}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s_f)^T \Delta q'_e(s_f) - \overline{C}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s_0)^T \Delta q'_e(s_0)\\
&+\overline{E}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s_f)\Delta\beta(s_f) - \overline{E}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s_0)\Delta\beta(s_0),
\end{aligned}$$

where

$$\overline{G}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s) := \overline{A}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s) - \overline{B}^{(0)'}_{g,\overline{q}_e,\overline{\beta}}(s) + \overline{C}^{(0)''}_{g,\overline{q}_e,\overline{\beta}}(s) \tag{10.39b}$$

$$\overline{H}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s) := \overline{D}^{(0)}_{g,\overline{q}_e,\overline{\beta}}(s) - \overline{E}^{(0)'}_{g,\overline{q}_e,\overline{\beta}}(s). \tag{10.39c}$$

Conditions (10.30e)–(10.30f), (10.30f'), (10.33c), (10.33d), resp., yield the following initial and terminal conditions for the changes $\Delta q_e(s)$, $\Delta\beta(s)$:

$$\Delta\beta(s_0) = 0,\ \Delta q_e(s_0) = \overline{q}_0 - \overline{q}_e(s_0) \tag{10.39d}$$

$$\Delta\beta(s_f) = 0,\ \Delta q_e(s_f) = \overline{q}_f - \overline{q}_e(s_f),\ \text{if } \phi^J = 0. \tag{10.39e}$$

Moreover, if $\dot{\overline{q}}_0 \neq 0$ (as in later correction stages, cf. Sect. 10.5), according to (10.30e) or (10.33c), condition $\Delta\beta(s_0) = 0$ must be replaced by the more general one

$$\left(\overline{q}'_e(s_0) + \Delta q'_e(s_0)\right)\sqrt{\overline{\beta}(s_0) + \Delta\beta(s_0)} = \dot{\overline{q}}_0 \tag{10.39f}$$

which can be approximated by

$$\sqrt{\overline{\beta}(s_0)}\,\Delta q'_e(s_0) + \frac{1}{2}\frac{\Delta\beta(s_0)}{\sqrt{\overline{\beta}(s_0)}}\overline{q}'_e(s_0) \approx \dot{\overline{q}}_0 - \sqrt{\overline{\beta}(s_0)}\,\overline{q}'_e(s_0). \tag{10.39f'}$$

Applying the above-described linearization (10.38a)–(10.38h) to (10.30a)–(10.30e) or to the constraints (10.30b), (10.30c) only, problem (10.30a)–(10.30f), (10.30f') is approximated by a linear variational problem or a variational problem having linear constraints for the changes $\Delta q_e(\cdot)$, $\Delta\beta(\cdot)$. On the other hand, using linearizations of the type (10.38a)–(10.38h) in the variational problem (10.33a)–(10.33f), in the average constraints (10.30b'), (10.30c'), resp., an optimization problem for $\Delta q_e(\cdot)$, $\Delta\beta(\cdot)$ is obtained which is linear, which has linear constraints, respectively.

(d) *Separated computation of* $q_e(\cdot)$ and $\beta(\cdot)$

In order to reduce the computational complexity, the given trajectory planning problem is often split up [16] into the following two separated problems for $q_e(\cdot)$ and $\beta(\cdot)$:

(i) Optimal path planning: find the shortest collision-free geometric path $q_e^{(0)} = q_e^{(0)}(s)$, $s_0 \le s \le s_f$, in configuration space from a given initial point q_0 to a prescribed terminal point q_f. Alternatively, with a given initial velocity profile $\beta(\cdot) = \overline{\beta}(\cdot)$, see (10.36b), the substitute problem (10.30a)–(10.30f), (10.30f'), (10.33a)–(10.33f), resp., may be solved for an approximate geometric path $q_e(\cdot) = q_e^{(0)}(\cdot)$ only.

(ii) Velocity planning: Determine then an optimal velocity profile $\beta^{(0)} = \beta^{(0)}(s)$, $s_0 \le s \le s_f$, along the predetermined path $q_e^{(0)}(\cdot)$.

Having a certain collection $\left\{q_{e,\lambda}(\cdot) : \lambda \in \Lambda\right\}$ of admissible paths in configuration space, a variant of the above procedure (i), (ii) is to determine—in an inner

optimization loop—the optimal velocity profile $\beta_\lambda(\cdot)$ with respect to a given path $q_{e,\lambda}(\cdot)$, and to optimize then the parameter λ in an outer optimization loop, see [19].

10.4.2 Optimal Reference Trajectory, Optimal Feedforward Control

Having, at least approximatively, the optimal geometric path $q_e^{(0)} = q_e^{(0)}(s)$ and the optimal velocity profile $\beta^{(0)} = \beta^{(0)}(s)$, $s_0 \le s \le s_f$, i.e., the optimal solution $\left(q_e^{(0)}, \beta^{(0)}\right) = \left(q_e^{(0)}(s), \beta^{(0)}(s)\right)$, $s_0 \le s \le s_f$, of one of the stochastic path planning problems (10.30a)–(10.30f), (10.30f'), (10.33a)–(10.33f), (10.34a)–(10.34d), resp., then, according to (10.13a), (10.13b), (10.14), the optimal reference trajectory in configuration space $q^{(0)} = q^{(0)}(t)$, $t \ge t_0$, is defined by

$$q^{(0)}(t) := q_e^{(0)}\left(s^{(0)}(t)\right), \quad t \ge t_0. \tag{10.40a}$$

Here, the optimal $(t \leftrightarrow s)$-transformation $s^{(0)} = s^{(0)}(t)$, $t \ge t_0$, is determined by the initial value problem

$$\dot{s}(t) = \sqrt{\beta^{(0)}(s)}, \quad t \ge t_0, \; s(t_0) := s_0. \tag{10.40b}$$

By means of the kinematic equation (10.5), the corresponding reference trajectory $x^{(0)} = x^{(0)}(t)$, $t \ge t_0$, in workspace may be defined by

$$x^{(0)}(t) := E\left(T\left(p_K(\omega), q^{(0)}(t)\right)|\mathcal{A}_{t_0}\right) = T\left(\overline{p}_K^{(0)}, q^{(0)}(t)\right), \quad t \ge t_0, \tag{10.40c}$$

where

$$\overline{p}_K^{(0)} := E\left(p_K(\omega)|\mathcal{A}_{t_0}\right). \tag{10.40d}$$

Based on the *inverse dynamics approach*, see Remark 10.1, the optimal reference trajectory $q^{(0)} = q^{(0)}(t)$, $t \ge t_0$, is inserted now into the left-hand side of the dynamic equation (10.4a). This yields next to the random optimal control function

$$\begin{aligned} v^{(0)}\left(t, p_D(\omega)\right) := {} & M\left(p_D(\omega), q^{(0)}(t)\right)\ddot{q}^{(0)}(t) \\ & + h\left(p_D(\omega), q^{(0)}(t), \dot{q}^{(0)}(t)\right), \quad t \ge t_0. \end{aligned} \tag{10.41}$$

Starting at the initial state $(\overline{q}_0, \overline{\dot{q}}_0) := \left(q^{(0)}(t_0), \dot{q}^{(0)}(t_0)\right)$, this control obviously keeps the robot exactly on the optimal trajectory $q^{(0)}(t)$, $t \ge t_0$, provided that $p_D(\omega)$ is the true vector of dynamic parameters.

An optimal feedforward control law $u^{(0)} = u^{(0)}(t), t \geq t_0$, related to the optimal reference trajectory $q^{(0)} = q^{(0)}(t), t \geq t_0$, can be obtained then by applying a certain averaging or estimating operator $\Psi = \Psi\left(\cdot|\mathcal{A}_{t_0}\right)$ to (10.41), hence,

$$u^{(0)} := \Psi\left(v^{(0)}\left(t, p_D(\cdot)\right)|\mathcal{A}_{t_0}\right), t \geq t_0. \tag{10.42}$$

If $\Psi(\cdot|\mathcal{A}_{t_0})$ is the conditional expectation, then we find the optimal feedforward control law

$$\begin{aligned} u^{(0)} &:= E\left(M\left(p_D(\omega), q^{(0)}(t)\right)\ddot{q}^{(0)}(t) + h\left(p_D(\omega), q^{(0)}(t), \dot{q}^{(0)}(t)\right)|\mathcal{A}_{t_0}\right), \\ &= M\left(\overline{p}_D^{(0)}, q^{(0)}(t)\right)\ddot{q}^{(0)}(t) + h\left(\overline{p}_D^{(0)}, q^{(0)}(t), \dot{q}^{(0)}(t)\right), \quad t \geq t_0, \end{aligned} \tag{10.43a}$$

where $\overline{p}_D^{(0)}$ denotes the conditional mean of $p_D(\omega)$ defined by (10.32c), and the second equation in formula (10.43a) holds since the dynamic equation of a robot depends linearly on the parameter vector p_D, see Remark 10.2.

Inserting into the dynamic equation (10.4a), instead of the conditional mean $\overline{p}_D^{(0)}$ of $p_D(\omega)$ given $\mathcal{A}_{t_0}$, another estimator $p_D^{(0)}$ of the true parameter vector p_D or a certain realization $p_D^{(0)}$ of $p_D(\omega)$ at the time instant t_0, then we obtain the optimal feedforward control law

$$u^{(0)}(t) := M\left(p_D^{(0)}, q^{(0)}(t)\right)\ddot{q}^{(0)}(t) + h\left(p_D^{(0)}, q^{(0)}(t), \dot{q}^{(0)}(t)\right), \quad t \geq t_0. \tag{10.43b}$$

10.5 AOSTP—Adaptive Optimal Stochastic Trajectory Planning

As already mentioned in the introduction, by means of direct or indirect measurements, observations of the robot and its environment, as, e.g., by observations of the state $(x, \dot{x})$, $(q, \dot{q})$, resp., of the mechanical system in work or configuration space, further information about the unknown parameter vector $p = p(\omega)$ is available at each moment $t > t_0$. Let denote, cf. [5],

$$\mathcal{A}_t (\subset \mathcal{A}), \quad t \geq t_0, \tag{10.44a}$$

the σ-algebra of all information about the random parameter vector $p = p(\omega)$ up to time t. Hence, $(\mathcal{A}_t)$ is an increasing family of σ-algebras. Note that the flow of information in this control process can be described also by means of the stochastic process

$$p_t(\omega) := E\left(p(\omega)|\mathcal{A}_t\right), \quad t \geq t_0, \tag{10.44b}$$

see [5].

By parameter identification [18, 43] or robot calibration techniques [6, 44] we may then determine the conditional distribution

$$P_{p(\cdot)}^{(t)} = P_{p(\cdot)|\mathcal{A}_t} \tag{10.44c}$$

of $p(\omega)$ given $\mathcal{A}_t$. Alternatively, we may determine the vector of conditional moments

$$\nu^{(t)} := \left(E\Big(\prod_{k=1}^{r} p_{l_k}(\omega) | \mathcal{A}_t \Big) \right)_{(l_1,\ldots,l_r)\in\Lambda} \tag{10.44d}$$

arising in the approximate computation of conditional expectations in (OSTP) with respect to $\mathcal{A}_t$, cf. (10.32c), (10.32d).

The increase of information about the unknown parameter vector $p(\omega)$ from one moment t to the next $t + dt$ may be rather low, and the determination of $P_{p(\cdot)}^{(t)}$ or $\nu^{(t)}$ at each time point t may be very expensive, though identification methods in real-time exist [43]. Hence, as already mentioned briefly in Sect. 10.1, the conditional distribution $P_{p(\cdot)}^{(t)}$ or the vector of conditional moments $\nu^{(t)}$ is determined/updated at discrete moments (t_j):

$$t_0 < t_1 < t_2 < \cdots < t_j < t_{j+1} < \ldots. \tag{10.45a}$$

The optimal functions $q_e^{(0)}(s), \beta^{(0)}(s), s_0 \le s \le s_f$, based on the a priori information $\mathcal{A}_{t_0}$, loose in course of time more or less their qualification to provide a satisfactory pair of guiding functions $\left(q^{(0)}(t), u^{(0)}(t)\right), t \ge t_0$.

However, having at the main correction time points t_j, $j = 1, 2, \ldots$, the updated information σ-algebras $\mathcal{A}_{t_j}$ and then the a posteriori probability distributions $P_{p(\cdot)}^{(t_j)}$ or the updated conditional moments $\nu^{(t_j)}$ of $p(\omega)$, $j = 1, 2, \ldots$, the pair of guiding functions $\left(q^{(0)}(t), u^{(0)}(t)\right), t \ge t_0$, is replaced by a sequence of renewed pairs $\left(q^{(j)}(t), u^{(j)}(t)\right), t \ge t_j$, $j = 1, 2, \ldots$, of guiding functions determined by replanning, i.e., by repeated (OSTP) for the remaining time intervals $[t_j, t_f^{(j)}]$ and by using the new information given by $\mathcal{A}_{t_j}$. Since replanning at a later main correction time point t_j, $j \ge 1$, hence on-line, may be very time consuming, in order to maintain the real-time capability of the method, we may start the replanning procedure for an update of the guiding functions at time t_j already at some earlier time $\tilde{t}_j$ with $t_{j-1} < \tilde{t}_j < t_j$, $j \ge 1$. Of course, in this case

$$\mathcal{A}_{t_j} := \mathcal{A}_{\tilde{t}_j} \tag{10.45b}$$

is defined to contain only the information about the control process up to time $\tilde{t}_j$ in which replanning, cf. Fig. 10.1, for time t_j starts.

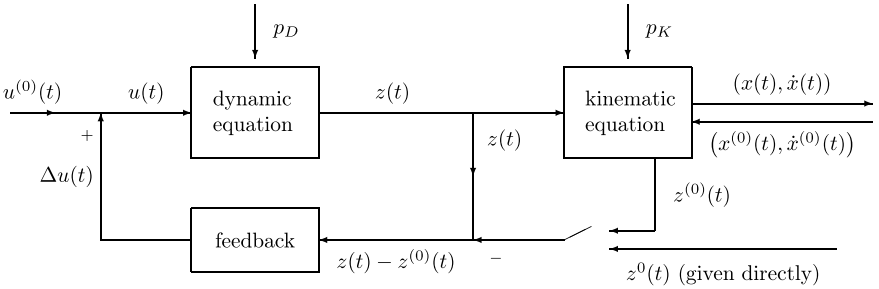


Fig. 10.1 Start of replanning

The resulting substitute problem at a stage $j \geq 1$ follows from the corresponding substitute problem for the previous stage $j-1$ just by updating $\zeta_{j-1} \to \zeta_j$, $\zeta_f^{(j-1)} \to \zeta_f^{(j)}$, the initial and terminal parameters, see (10.32a), (10.32b). The renewed

$$\text{initial parameters } \zeta_j : \ t_j, s_j, \overline{q}_j, \overline{\dot{q}}_j, P_{p(\cdot)}^{(j)} \text{ or } \nu_j \tag{10.46a}$$

for the j-th stage, $j \geq 1$, are determined recursively as follows:

$$s_j := s^{(j-1)}(t_j) \quad (1-1-\text{transformation } s = s(t)) \tag{10.46b}$$

$$\overline{q}_j := \hat{q}(t_j), \overline{q}_j := q^{(j-1)}(t_j) \text{ or } \overline{q}_j := E\Big(q(t_j)|\mathcal{A}_{t_j}\Big) \tag{10.46c}$$

$$\overline{\dot{q}}_j := \hat{\dot{q}}(t_j), \overline{q}_j := \dot{q}^{(j-1)}(t_j) \text{ or } \overline{\dot{q}}_j := E\Big(\dot{q}(t_j)|\mathcal{A}_{t_j}\Big) \tag{10.46d}$$

(observation or estimate of $q(t_j), \dot{q}(t_j)$)

$$P_{p(\cdot)}^{(j)} := P_{p(\cdot)}^{(t_j)} = P_{p(\cdot)|\mathcal{A}_{t_j}} \tag{10.46e}$$

$$\nu_j := \nu^{(t_j)}. \tag{10.46f}$$

The renewed

$$\text{terminal parameters } \zeta_f^{(j)} : \ t_f^{(j)}, s_f, \overline{q}_f^{(j)}, \beta_f \tag{10.47a}$$

for the j-th stage, $j \geq 1$, are defined by

$$s_f \text{ (given)} \tag{10.47b}$$

$$\overline{q}_f^{(j)} := \hat{q}(t_f) \text{ or } \overline{q}_f^{(j)} := E\Big(q_f(\omega)|\mathcal{A}_{t_j}\Big) \tag{10.47c}$$

$$\beta_f = 0 \tag{10.47d}$$

$$s^{(j)}\left(t_f^{(j)}\right) = s_f. \tag{10.47e}$$

As already mentioned above, the (OSTP) for the j-th stage, $j \geq 1$, is obtained from the substitute problems (10.30a)–(10.30f), (10.30f'), (10.33a)–(10.33f), (10.34a)–(10.34d), resp., formulated for the 0-th stage, $j = 0$, just by substituting

$$\zeta_0 \to \zeta_j \text{ and } \zeta_f \to \zeta_f^{(j)}. \tag{10.48}$$

Let then denote

$$\left(q_e^{(j)}, \beta^{(j)}\right) = \left(q_e^{(j)}(s), \beta^{(j)}(s)\right), \quad s_j \leq s \leq s_f, \tag{10.49}$$

the corresponding pair of optimal solutions of the resulting substitute problem for the j-th stage, $j \geq 1$.

The pair of guiding functions $\left(q^{(j)}(t), u^{(j)}(t)\right), t \geq t_j$, for the j-th stage, $j \geq 1$, is then defined as described in Sect. 10.4.2 for the 0-th stage. Hence, for the j-th stage, the reference trajectory in configuration space $q^{(j)}(t), t \geq t_j$, reads cf. (10.40a),

$$q^{(j)}(t) := q_e^{(j)}\left(s^{(j)}(t)\right), \quad t \geq t_j, \tag{10.50a}$$

where the transformation $s^{(j)} : \left[t_j, t_f^{(j)}\right] \to [s_j, s_f]$ is defined by the initial value problem

$$\dot{s}(t) = \sqrt{\beta^{(j)}(s)}, \quad t \geq t_j, \quad s(t_j) = s_j. \tag{10.50b}$$

The terminal time $t_f^{(j)}$ for the j-th stage is defined by the equation

$$s^{(j)}\left(t_f^{(j)}\right) = s_f. \tag{10.50c}$$

Moreover, again by the inverse dynamics approach, the feedforward control $u^{(j)} = u^{(j)}(t), t \geq t_j$, for the j-th stage is defined, see (10.41), (10.42), (10.43a), (10.43b), by

$$u^{(j)}(t) := \Psi\left(v^{(j)}\left(t, p_D(\omega)\right)|\mathcal{A}_{t_j}\right), \tag{10.51a}$$

where

$$v^{(j)}(t, p_D) := M\left(p_D, q^{(j)}(t)\right)\ddot{q}^{(j)}(t) + h\left(p_D, q^{(j)}(t), \dot{q}^{(j)}(t)\right), \quad t \geq t_j. \tag{10.51b}$$

Using the conditional expectation $\Psi(\cdot|\mathcal{A}_{t_j}) := E(\cdot|\mathcal{A}_{t_j})$, we find the feedforward control

$$u^{(j)}(t) := M\left(\overline{p}_D^{(j)}, q^{(j)}(t)\right)\ddot{q}^{(j)}(t) + h\left(\overline{p}_D^{(j)}, q^{(j)}, \dot{q}^{(j)}(t)\right), t \geq t_j, \tag{10.51c}$$

where, cf. (10.32c),

$$\overline{p}_D^{(j)} := E\left(p_D(\omega)|\mathcal{A}_{t_j}\right). \tag{10.51d}$$

Corresponding to (10.40c), (10.40d), the reference trajectory in work space $x^{(j)} = x^{(j)}(t), t \geq t_j$, for the remaining time interval $t_j \leq t \leq t_f^{(j)}$, is defined by

$$x^{(j)}(t) := E\left(T\left(p_K(\omega), q^{(j)}(t)\right)|\mathcal{A}_{t_j}\right) = T\left(\overline{p}_K^{(j)}, q^{(j)}(t)\right), \quad t_j \leq t \leq t_f^{(j)}, \tag{10.52a}$$

where

$$\overline{p}_K^{(j)} := E\left(p_K(\omega)|\mathcal{A}_{t_j}\right). \tag{10.52b}$$

10.5.1 (OSTP)-Transformation

The variational problems (OSTP) at the different stages $j = 0, 1, 2 \ldots$ are determined uniquely by the set of initial and terminal parameters $(\zeta_j, \zeta_f^{(j)})$, cf. (10.46a)–(10.46f), (10.47a)–(10.47e). Thus, these problems can be transformed to a reference problem depending on $\left(\zeta_j, \zeta_f^{(j)}\right)$ and having a certain fixed reference s-interval.

Theorem 10.1 *Let $[\tilde{s}_0, \tilde{s}_f], \tilde{s}_0 < \tilde{s}_f := s_f$, be a given, fixed reference s-interval, and consider for a certain stage j, $j = 0, 1, \ldots$, the transformation*

$$\tilde{s} = \tilde{s}(s) := \tilde{s}_0 + \frac{\tilde{s}_f - \tilde{s}_0}{s_f - s_j}(s - s_j), \quad s_j \leq s \leq s_f, \tag{10.53a}$$

from $[s_j, s_f]$ onto $[\tilde{s}_0, s_f]$ having the inverse

$$s = s(\tilde{s}) = s_j + \frac{s_f - s_j}{\tilde{s}_f - \tilde{s}_0}(\tilde{s} - \tilde{s}_0), \quad \tilde{s}_0 \leq \tilde{s} \leq \tilde{s}_f. \tag{10.53b}$$

Represent then the geometric path in work space $q_e = q_e(s)$ and the velocity profile $\beta = \beta(s), s_j \leq s \leq s_f$, for the j-th stage by

$$q_e(s) := \tilde{q}_e\left(\tilde{s}(s)\right), \quad s_j \leq s \leq s_f \tag{10.54a}$$

$$\beta(s) := \tilde{\beta}\left(\tilde{s}(s)\right), \quad s_j \leq s \leq s_f, \tag{10.54b}$$

where $\tilde{q}_e = \tilde{q}_e(\tilde{s}), \tilde{\beta} = \tilde{\beta}(\tilde{s}), \tilde{s}_0 \leq \tilde{s} \leq \tilde{s}_f$, denote the corresponding functions on $[\tilde{s}_0, \tilde{s}_f]$. Then the (OSTP) for the j-th stage is transformed into a reference variational problem (stated in the following) for $(\tilde{q}_e, \tilde{\beta})$ depending on the parameters

$$(\zeta, \zeta_f) = (\zeta_j, \zeta_f^{(j)}) \in Z \times Z_f \tag{10.55}$$

and having the fixed reference s-interval $[\tilde{s}_0, \tilde{s}_f]$. Moreover, the optimal solution $\left(q_e^{(j)}, \beta^{(j)}\right) = \left(q_e^{(j)}(s), \beta^{(j)}(s)\right), s_j \le s \le s_f$, may be represented by the optimal adaptive law

$$q_e^{(j)}(s) = \tilde{q}_e^*\left(\tilde{s}(s); \zeta_j, \zeta_f^{(j)}\right), \quad s_j \le s \le s_f, \tag{10.56a}$$

$$\beta^{(j)}(s) = \tilde{\beta}^*\left(\tilde{s}(s); \zeta_j, \zeta_f^{(j)}\right), \quad s_j \le s \le s_f, \tag{10.56b}$$

where

$$\tilde{q}_e^* = \tilde{q}_e^*(\tilde{s}; \zeta, \zeta_f), \quad \tilde{\beta}^* = \tilde{\beta}^*(\tilde{s}; \zeta, \zeta_f), \quad \tilde{s}_0 \le \tilde{s} \le \tilde{s}_f \tag{10.56c}$$

denotes the optimal solution of the above-mentioned reference variational problem.

Proof According to (10.54a), (10.54b) and (10.53a), (10.53b), the derivatives of the functions $q_e(s), \beta(s), s_j \le s \le s_f$, are given by

$$q_e'(s) = \tilde{q}_e'\left(\tilde{s}(s)\right)\frac{\tilde{s}_f - \tilde{s}_0}{s_f - s_j}, \quad s_j \le s \le s_f, \tag{10.57a}$$

$$q_e''(s) = \tilde{q}_e''\left(\tilde{s}(s)\right)\left(\frac{\tilde{s}_f - \tilde{s}_0}{s_f - s_j}\right)^2, \quad s_j \le s \le s_f, \tag{10.57b}$$

$$\beta'(s) = \tilde{\beta}'\left(\tilde{s}(s)\right)\frac{\tilde{s}_f - \tilde{s}_0}{s_f - s_j}, \quad s_j \le s \le s_f. \tag{10.57c}$$

Now putting the transformation (10.53a), (10.53b) and the representation (10.54a), (10.54b), (10.57a)–(10.57c) of $q_e(x), \beta(s), s_j \le s \le s_f$, and their derivatives into one of the substitute problems (10.30a)–(10.30f), (10.30f'), (10.33a)–(10.33f) or their time-variant versions, the chosen substitute problem is transformed into a corresponding reference variational problem (stated in the following Sect. 10.5.2) having the fixed reference interval $[\tilde{s}_0, \tilde{s}_f]$ and depending on the parameter vectors $\zeta_j, \zeta_f^{(j)}$. Moreover, according to (10.54a), (10.54b), the optimal solution $\left(q_e^{(j)}, \beta^{(j)}\right)$ of the substitute problem for the j-th stage may be represented then by (10.56a)–(10.56c). □

Remark 10.11 Based on the above theorem, the stage-independent functions $\tilde{q}_e^*, \tilde{\beta}^*$ can be determined now offline by using an appropriate numerical procedure.

10.5.2 The Reference Variational Problem

After the (OSTP)-transformation described in Sect. 10.5.1, in the *time- invariant case* for the problems of type (10.30a)–(10.30f), (10.30f') we find

$$\min \int_{\tilde{s}_0}^{\tilde{s}_f} E\left(L^J\Big(p_J, \tilde{q}_e(\tilde{s}), \tilde{q}_e'(\tilde{s})\frac{\tilde{s}_f-\tilde{s}_0}{s_f-s_j}, \tilde{q}_e''(\tilde{s})\left(\frac{\tilde{s}_f-\tilde{s}_0}{s_f-s_j}\right)^2, \tilde{\beta}(\tilde{s}), \tilde{\beta}'(\tilde{s})\frac{\tilde{s}_f-\tilde{s}_0}{s_f-s_j}\Big)|\mathcal{A}_{t_j}\right)\frac{s_f-s_j}{\tilde{s}_f-\tilde{s}_0}\,d\tilde{s} + E\left(\phi^J\Big(p_J, \tilde{q}_e(\tilde{s_f})\Big)|\mathcal{A}_{t_j}\right) \quad (10.58a)$$

s.t.

$$E\left(f_\gamma\Big(p, \tilde{q}_e(\tilde{s}), \tilde{q}_e'(\tilde{s})\frac{\tilde{s}_f-\tilde{s}_0}{s_f-s_j}, \tilde{q}_e''(\tilde{s})\left(\frac{\tilde{s}_f-\tilde{s}_0}{s_f-s_j}\right)^2, \tilde{\beta}(\tilde{s}), \tilde{\beta}'(\tilde{s})\frac{\tilde{s}_f-\tilde{s}_0}{s_f-s_j}\Big)|\mathcal{A}_{t_j}\right) \le \Gamma_f, \quad (10.58b)$$

$$\tilde{\beta}(\tilde{s}) \ge 0, \quad \tilde{s}_0 \le \tilde{s} \le \tilde{s}_f \quad (10.58c)$$

$$\tilde{q}_e(\tilde{s}_0) = \overline{q}_j, \quad \tilde{q}_e'(\tilde{s}_0)\frac{\tilde{s}_f-\tilde{s}_0}{s_f-s_j}\sqrt{\tilde{\beta}(\tilde{s}_0)} = \overline{\dot{q}}_j \quad (10.58d)$$

$$\tilde{q}_e(\tilde{s}_f) = \overline{q}_f^{(j)} \quad (\text{if } \phi^J = 0), \quad \tilde{\beta}(\tilde{s}_f) = 0 \quad (10.58e)$$

$$\tilde{\beta}(\tilde{s}_f) = 0, \quad E\left(\psi\Big(p, \tilde{q}_e(\tilde{s}_f)\Big)|\mathcal{A}_{t_j}\right) \le \Gamma_\psi, \quad (10.58e')$$

where f_γ, Γ_f are defined by

$$f_\gamma := \begin{pmatrix} f_\gamma^u \\ f_\gamma^S \end{pmatrix}, \quad \Gamma_f := \begin{pmatrix} \Gamma_f^u \\ \Gamma_f^S \end{pmatrix}. \quad (10.58f)$$

Moreover, for the problem type (10.33a)–(10.33f) we get

$$\min \int_{\tilde{s}_0}^{\tilde{s}_f} E\left(L_\gamma^J\Big(p, \tilde{q}_e(\tilde{s}), \tilde{q}_e'(\tilde{s})\frac{\tilde{s}_f-\tilde{s}_0}{s_f-s_j}, \tilde{q}_e''(\tilde{s})\left(\frac{\tilde{s}_f-\tilde{s}_0}{s_f-s_j}\right)^2, \tilde{\beta}(\tilde{s}), \tilde{\beta}'(\tilde{s})\frac{\tilde{s}_f-\tilde{s}_0}{s_f-s_j}\Big)|\mathcal{A}_{t_j}\right)\frac{s_f-s_j}{\tilde{s}_f-\tilde{s}_0}\,d\tilde{s} + E\left(\phi_\gamma^J\Big(p, \tilde{q}_e(\tilde{s}_f)\Big)|\mathcal{A}_{t_j}\right) \quad (10.59a)$$

s.t.

$$\tilde{\beta}(\tilde{s}) \ge 0, \quad \tilde{s}_0 \le \tilde{s} \le \tilde{s}_f \quad (10.59b)$$

$$\tilde{q}_e(\tilde{s}_0) = \overline{q}_j, \quad \tilde{q}_e'(\tilde{s}_0)\frac{\tilde{s}_f-\tilde{s}_0}{s_f-s_j}\sqrt{\tilde{\beta}(\tilde{s}_0)} = \overline{\dot{q}}_j \quad (10.59c)$$

$$\tilde{q}_e(\tilde{s}_f) = \overline{q}_f^{(j)} \quad (\text{ if } \phi_\gamma^J = 0), \quad \tilde{\beta}(\tilde{s}_f) = 0. \quad (10.59d)$$

For the consideration of the *time-variant case* we note first that by using the transformation (10.53a), (10.53b) and (10.54b) the time $t \ge t_j$ can be represented, cf. (10.15a), (10.15b) and (10.16a), also by

$$t = t\Big(\tilde{s}, t_j, s_j, \tilde{\beta}(\cdot)\Big) := t_j + \frac{s_f - s_j}{\tilde{s}_f - \tilde{s}_0} \int_{\tilde{s}_0}^{\tilde{s}} \frac{d\tilde{\sigma}}{\sqrt{\tilde{\beta}(\tilde{\sigma})}}. \tag{10.60a}$$

Hence, if the variational problems (10.58a)–(10.58f) and (10.59)–(10.59d) for the j-th stage depend explicitly on time $t \geq t_j$, then, corresponding to Sect. 10.4, item *(B)*, for the constituting functions $L^J, \phi^J, L^J_\gamma, \phi^J_\gamma$ of the variational problems we have that

$$L^J = L^J\Big(\tilde{s}, t_j, s_j, \tilde{\beta}(\cdot), p_J, q_e, q'_e, q''_e, \beta, \beta'\Big), \quad \tilde{s}_0 \leq \tilde{s} \leq \tilde{s}_f \tag{10.60b}$$

$$\phi^J = \phi^J\Big(\tilde{s}_f, t_j, s_j, \tilde{\beta}(\cdot), p_J, q_e\Big) \tag{10.60c}$$

$$f_\gamma = f_\gamma\Big(\tilde{s}, t_j, s_j, \tilde{\beta}(\cdot), p, q_e, q'_e, q''_e, \beta, \beta'\Big), \quad \tilde{s}_0 \leq \tilde{s} \leq \tilde{s}_f \tag{10.60d}$$

$$L^J_\gamma = L^J_\gamma\Big(\tilde{s}, t_j, s_j, \tilde{\beta}(\cdot), p, q_e, q''_e, \beta, \beta'\Big), \quad \tilde{s}_0 \leq \tilde{s} \leq \tilde{s}_f \tag{10.60e}$$

$$\phi^J_\gamma = \phi^J_\gamma\Big(\tilde{s}_f, t_j, s_j, \tilde{\beta}(\cdot), p, q_e\Big). \tag{10.60f}$$

10.5.2.1 Transformation of the Initial State Values

Suppose here that $\phi^J \neq 0, \phi^J_\gamma \neq 0$, resp., and the terminal state condition (10.58e), (10.58e'), (10.59d), resp., is reduced to

$$\tilde{\beta}(\tilde{s}_f) = 0. \tag{10.61a}$$

Representing then the unknown functions $\tilde{\beta}(\cdot), \tilde{q}_e(\cdot)$ on $[\tilde{s}_0, \tilde{s}_f]$ by

$$\tilde{\beta}(\tilde{s}) := \beta_j \tilde{\beta}_a(\tilde{s}), \quad \tilde{s}_0 \leq \tilde{s} \leq \tilde{s}_f \tag{10.61b}$$

$$\tilde{q}_e(\tilde{s}) := \overline{q}_{jd} \tilde{q}_{ea}(\tilde{s}), \quad \tilde{s}_0 \leq \tilde{s} \leq \tilde{s}_f, \tag{10.61c}$$

where $\overline{q}_{jd}$ denotes the diagonal matrix with the components of $\overline{q}_j$ on its main diagonal, then in terms of the new unknowns $\Big(\tilde{\beta}_a(\cdot), \tilde{q}_{ea}(\cdot)\Big)$ on $[\tilde{s}_0, \tilde{s}_f]$ we have the nonnegativity and fixed initial/terminal conditions

$$\tilde{\beta}_a(\tilde{s}) \geq 0, \quad \tilde{s}_0 \leq \tilde{s} \leq \tilde{s}_f \tag{10.62a}$$

$$\tilde{\beta}_a(\tilde{s}_0) = 1, \quad \tilde{q}_{ea}(\tilde{s}_0) = \mathbf{1} \tag{10.62b}$$

$$\tilde{\beta}_a(\tilde{s}_f) = 0, \tag{10.62c}$$

where $\mathbf{1} := (1, \ldots, 1)'$.

10.5.3 Numerical Solutions of (OSTP) in Real-Time

With the exception of field robots (e.g., Mars rover) and service robots [16], becoming increasingly important, the standard industrial robots move very fast. Hence, for industrial robots the optimal solution $\left(q_e^{(j)}(s), \beta^{(j)}(s)\right)$, $\beta^{(j)}(s)$, resp., $s_j \le s \le s_f$, generating the renewed pair of guiding functions $\left(q^{(j)}(t), u^{(j)}(t)\right)$, $t \ge t_j$, on each stage $j = 1, 2, \ldots$ should be provided in *real-time*. This means that the optimal solutions $\left(q_e^{(j)}, \beta^{(j)}\right)$, $\beta^{(j)}$, resp., must be prepared offline as far as possible such that only relatively simple numerical operations are left online.

Numerical methods capable to generate approximate optimal solutions in real-time are based mostly on discretization techniques, neural network (NN) approximation [3, 30, 31, 38], linearization techniques (sensitivity analysis) [48].

10.5.3.1 Discretization Techniques

Partitioning the space $Z \times Z_f$ of initial/terminal parameters (ζ, ζ_f) into a certain (small) number l_0 of subdomains

$$Z \times Z_f = \bigcup_{l=1}^{l_0} Z^l \times Z_f^l, \tag{10.63a}$$

and selecting then a reference parameter vector

$$(\zeta^l, \zeta_f^l) \in Z^l \times Z_f^l, l = 1, \ldots, l_0, \tag{10.63b}$$

in each subdomain $Z^l \times Z_f^l$, the optimal adaptive law (10.56c) can be approximated, cf. [47], by

$$\left.\begin{aligned} \hat{\tilde{q}}_e^*(\tilde{s}; \zeta, \zeta_f) &:= \tilde{q}^*(\tilde{s}; \zeta^l, \zeta_f^l), \tilde{s}_0 \le \tilde{s} \le \tilde{s}_f \\ \hat{\tilde{\beta}}^*(\tilde{s}; \zeta, \zeta_f) &:= \tilde{\beta}^*(\tilde{s}; \zeta^l, \zeta_f^l), \tilde{s}_0 \le \tilde{s} \le \tilde{s}_f \end{aligned}\right\} \text{ for } (\zeta, \zeta_f) \in Z^l \times Z_f^l. \tag{10.63c}$$

10.5.3.2 NN-Approximation

For the determination of the optimal adaptive law (10.56a)–(10.56c) in real-time, according to (10.35a), (10.35b), the reference variational problem (10.58a)–(10.58f) or (10.59)–(10.59d) is reduced first to a finite-dimensional parameter optimization problem by

(i) representing the unknown functions $\tilde{q}_e = \tilde{q}_e(\tilde{s})$, $\tilde{\beta} = \tilde{\beta}(\tilde{s})$, $\tilde{s}_0 \le \tilde{s} \le \tilde{s}_f$, as linear combinations

$$\tilde{q}_e(\tilde{s}) := \sum_{l=1}^{l_q} \hat{q}_e B_l^q(\tilde{s}),\ \tilde{s}_0 \le \tilde{s} \le \tilde{s}_f, \tag{10.64a}$$

$$\tilde{\beta}(\tilde{s} := \sum_{l=1}^{l_\beta} \hat{\beta}_l B_l^\beta(\tilde{s}),\ \tilde{s}_0 \le \tilde{s} \le \tilde{s}_f, \tag{10.64b}$$

of certain basis functions, e.g., cubic B-splines, $B_l^q = B_l^q(\tilde{s})$, $B_l^\beta = B_l^\beta(\tilde{s})$, $\tilde{s}_0 \le \tilde{s} \le \tilde{s}_f$, $l = 1, \ldots, l_q(l_\beta)$, with unknown vectorial (scalar) coefficients $\hat{q}_l$, $\hat{\beta}_l$, $l = 1, \ldots, l_q(l_\beta)$, and

(ii) demanding the inequalities in (10.58b), (10.58c), (10.59b), resp., only for a finite set of $\tilde{s}$-parameter values $\tilde{s}_0 < \tilde{s}_1 < \cdots < \tilde{s}_k < \tilde{s}_{k+1} < \cdots < \tilde{s}_\kappa = \tilde{s}_f$.

By means of the above-described procedure (i), (ii), the optimal coefficients

$$\hat{q}_l^* = \hat{q}_l^*(\zeta, \zeta_f), \quad l = 1, \ldots, l_q \tag{10.64c}$$

$$\hat{\beta}_l^* = \hat{\beta}_l^*(\zeta, \zeta_f), \quad l = 1, \ldots, l_\beta \tag{10.64d}$$

become functions of the initial/terminal parameters ζ, ζ_f, cf. (10.56c). Now, for the numerical realization of the optimal parameter functions (10.64c), (10.64d), a Neural Network (NN) is employed generating an approximative representation

$$\hat{q}_e^*(\zeta, \zeta_f) \approx \hat{q}_e^{NN}(\zeta, \zeta_f; w_q), \quad l = 1, \ldots, l_q \tag{10.65a}$$

$$\hat{\beta}_l^*(\zeta, \zeta_f) \approx \hat{\beta}_l^{NN}(\zeta, \zeta_f; w_\beta), \quad l = 1, \ldots, l_\beta, \tag{10.65b}$$

where the vectors w_q, w_β of NN-weights are determined optimally

$$w_q = w_q^* \text{ (data)}, \ w_\beta = w_\beta^* \text{ (data)} \tag{10.65c}$$

in an offline training procedure [3, 30, 38]. Here, the model (10.65a), (10.65b) is fitted in the LSQ-sense to given data

$$\begin{pmatrix} \zeta^\tau, \zeta_f^\tau \\ \hat{q}_l^{*\tau}, l = 1, \ldots, l_q \end{pmatrix}, \begin{pmatrix} \zeta^\tau, \zeta_f^\tau \\ \hat{\beta}_l^{*\tau}, l = 1, \ldots, l_\beta \end{pmatrix}, \quad \tau = 1, \ldots, \tau_0, \tag{10.65d}$$

where

$$(\zeta^\tau, \zeta_f^\tau), \quad \tau = 1, \ldots, \tau_0 \tag{10.65e}$$

is a certain collection of initial/terminal parameter vectors, and

$$\hat{q}_l^{*\tau} := \hat{q}_l^{*\tau}(\zeta^\tau, \zeta_f^\tau), \quad l = 1, \ldots, l_q, \quad \tau = 1, \ldots, \tau_0 \tag{10.65f}$$

$$\hat{\beta}_l^{*\tau} := \hat{\beta}_l^{*\tau}(\zeta^\tau, \zeta_f^\tau), \quad l = 1, \ldots, l_\beta, \quad \tau = 1, \ldots, \tau_0 \tag{10.65g}$$

are the related optimal coefficients in (10.64a), (10.64b) which are determined offline by an appropriate parameter optimization procedure.

Having the vectors w_q^*, w_β^* of optimal NN-weights, by means of (10.65a)–(10.65c), for given actual initial/terminal parameters $(\zeta, \zeta_f) = (\zeta_j, \zeta_f^{(j)})$ at stage $j \geq 0$, the NN yields then the optimal parameters

$$\hat{q}_l^*(\zeta_j, \zeta_f^{(j)}), \hat{\beta}_l^*(\zeta_j, \zeta_f^{(j)}), \quad l = 1, \ldots, l_q(l_\beta)$$

in real-time; consequently, by means of (10.64a), (10.64b), also the optimal functions $\tilde{q}_e^*(\tilde{s}), \tilde{\beta}^*(\tilde{s}), \tilde{s}_0 \leq \tilde{s} \leq \tilde{s}_f$, are then available very fast. For more details, see [3, 30].

10.5.3.3 Linearization Methods

(I) *Linearization of the optimal feedforward control law*
Expanding the optimal control laws (10.56c) with respect to the initial/terminal parameter vector $\hat{\zeta} = (\zeta, \zeta_f)$ at its value $\hat{\zeta}_0 = (\zeta_0, \zeta_f^{(0)})$ for stage $j = 0$, approximatively we have that

$$\tilde{q}_e^*(\tilde{s}, \zeta, \zeta_f) = \tilde{q}_e^*(\tilde{s}; \zeta_0, \zeta_f^{(0)}) + \frac{\partial \tilde{q}_e^*}{\partial \hat{\zeta}}(\tilde{s}; \zeta_0, \zeta_f^{(0)})(\zeta - \zeta_0, \zeta_f - \zeta_f^{(0)}) + \ldots \tag{10.66a}$$

$$\tilde{\beta}^*(\tilde{s}; \zeta, \zeta_f) = \tilde{\beta}^*(\tilde{s}; \zeta_0, \zeta_f^{(0)}) + \frac{\partial \tilde{\beta}^*}{\partial \hat{\zeta}}(\tilde{s}; \zeta_0, \zeta_f^{(0)})(\zeta - \zeta_0, \zeta_f - \zeta_f^{(0)}) + \ldots, \tag{10.66b}$$

where the optimal starting functions $\tilde{q}_e^*(\tilde{s}; \zeta_0, \zeta_f^{(0)}), \tilde{\beta}^*(\tilde{s}; \zeta_0, \zeta_f^{(0)})$ and the derivatives $\frac{\partial \tilde{q}_e^*}{\partial \hat{\zeta}}(\tilde{s}; \zeta_0, \zeta_f^{(0)}), \frac{\partial \tilde{\beta}^*}{\partial \hat{\zeta}}(\tilde{s}; \zeta_0, \zeta_f^{(0)}), \ldots$ can be determined—on a certain grid of $[\tilde{s}_0, \tilde{s}_f]$—offline by using sensitivity analysis [48]. The actual values of $\tilde{q}_e^*, \tilde{\beta}^*$ at later stages can then be obtained very rapidly by means of simple matrix operations. If necessary, the derivatives can be updated later on by a numerical procedure running in parallel to the control process.

(II) *Sequential linearization of the (AOSTP) process*
Given the optimal guiding functions $q_e^{(j)} = q_e^{(j)}(s), \beta^{(j)} = \beta^{(j)}(s), s_j \leq s \leq s_f$ for the j-th stage, corresponding to the representation (10.36a), (10.36b), the optimal guiding functions $q_e^{(j+1)}(s), \beta^{(j+1)}(s), \ s_{j+1} \leq s \leq s_f$, are represented, cf. Fig. 10.2, by

$$q_e^{(j+1)}(s) := q_e^{(j)}(s) + \Delta q_e(s), \quad s_{j+1} \leq s \leq s_f, \tag{10.67a}$$

$$\beta^{(j+1)}(s) := \beta^{(j)}(s) + \Delta\beta(s), \quad s_{j+1} \leq s \leq s_f, \tag{10.67b}$$

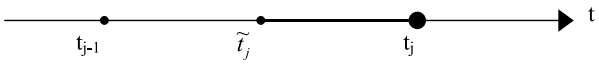


Fig. 10.2 Sequential linearization of (AOSTP)

where $s_j < s_{j+1} < s_f$, and $\left(\Delta q_e(s), \Delta\beta(s)\right)$, $s_{j+1} \le s \le s_f$, are certain (small) changes of the j-th stage optimal guiding functions $\left(q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\right)$.
Obviously, the linearization technique described in Sect. 10.5.3.3 can be applied now also to the approximate computation of the optimal changes $\Delta q_e(s)$, $\Delta\beta(s)$, $s_{j+1} \le s \le s_f$, if the following replacements are made in the formulas (10.36a), (10.36b), (10.37a)–(10.37c), (10.38a)–(10.38h) and (10.39a)–(10.39f):

$$\left.\begin{array}{ll}
s_0 \le s \le s_f & \to s_{j+1} \le s \le s_f \\
\overline{q}_e, \overline{\beta} & \to q_e^{(j)}, \beta^{(j)} \\
\mathcal{A}_{t_0} & \to \mathcal{A}_{t_{j+1}} \\
\overline{g}^{(0)}, \overline{\phi}^{(0)}, \overline{F}^{(0)} & \to \overline{g}^{(j+1)}, \overline{\phi}^{(j+1)}, \overline{F}^{(j+1)} \\
\overline{A}^{(0)}_{g,\overline{q}_e,\overline{\beta}}, \ldots, \overline{H}^{(0)}_{g,\overline{q}_e,\overline{\beta}} & \to \overline{A}^{(j+1)}_{g,q_e^{(j)},\beta^{(j)}}, \ldots, \overline{H}^{(j+1)}_{g,q_e^{(j)},\beta^{(j)}} \\
\Delta\beta(s_0) = 0 & \to \Delta\beta(s_{j+1}) = \beta_{j+1} - \beta^{(j)}(s_{j+1}) \\
\Delta q_e(s_0) = q_0 - \overline{q}_e(s_0) & \to \Delta q_e(s_{j+1}) = \overline{q}_{j+1} - q_e^{(j)}(s_{j+1}) \\
\Delta\beta(s_f) = 0 & \to \Delta\beta(s_f) = 0 \\
\Delta q_e(s_f) = q_f^{(0)} - \overline{q}_e(s_f) & \to \Delta q_e(s_f) = \overline{q}_f^{(j+1)} - \overline{q}_f^{(j)}, \text{ if } \phi^J = 0,
\end{array}\right\} \tag{10.67c}$$

where $\overline{X}^{(j+1)}$ denotes the conditional expectation of a random variable X with respect to $\mathcal{A}_{t_{j+1}}$, cf. (10.51d), (10.52b). Furthermore, (10.38f) yields

$$\left(q_e^{(j)'}(s_{j+1}) + \Delta q_e'(s_{j+1})\right) \cdot \sqrt{\beta_{j+1}} = \overline{\dot{q}}_{j+1} \tag{10.67d}$$

which can be approximated, cf. (10.39f'), by

$$\begin{aligned}
&\sqrt{\beta^{(j)}(s_{j+1})}\Delta q_e'(s_{j+1}) + \tfrac{1}{2}\frac{\Delta\beta(s_{j+1})}{\sqrt{\beta^{(j)}(s_{j+1})}} q_e^{(j)'}(s_{j+1}) \\
&\qquad \approx \overline{\dot{q}}_{j+1} - \sqrt{\beta^{(j)}(s_{j+1})} q_e^{(j)'}(s_{j+1}).
\end{aligned} \tag{10.67d'}$$

Depending on the chosen substitute problem, by this linearization method we obtain then a variational problem, an optimization problem, resp., for the changes $\left(\Delta q_e(s), \Delta\beta(s)\right)$, $s_{j+1} \le s \le s_f$, having a linear objective function and/or linear constraints.

To give a typical example, we consider now (AOSTP) on the $(j+1)$th stage with substitute problem (10.33a)–(10.33f). Hence, the functions g, ϕ in (10.40a)–(10.40c), (10.38a)–(10.38h), and (10.39a)–(10.39f) are given by

$$g := L_\gamma^J, \phi := \phi_\gamma^J.$$

Applying the linearization techniques developed in Sect. 10.4.1c now to (10.33a)–(10.33f), according to (10.39a)–(10.39c), (10.38b) and (10.67d'), for the correction terms $\Delta q_e(s)$, $\Delta\beta(s)$, $s_{j+1} \le s \le s_f$, we find the following linear optimization problem:

$$\min \int_{s_{j+1}}^{s_f} \overline{G}_{g,q_e^{(j)},\beta^{(j)}}^{(j+1)}(s)^T \Delta q_e(s) ds + \int_{s_{j+1}}^{s_f} \overline{H}_{g,q_e^{(j)},\beta^{(j)}}^{(j+1)}(s) \Delta(s) ds \tag{10.68a}$$
$$+ R_j^T \Delta q_e(s_f) + S_j^T \Delta q_e'(s_f) + T_j \Delta\beta(s_{j+1})$$

s.t.

$$\Delta q_e(s_{j+1}) = \overline{q}_{j+1} - q_e^{(j)}(s_{j+1}) \tag{10.68b}$$
$$\Delta q_e(s_f) = \overline{q}_f^{(j+1)} - \overline{q}_f^{(j)}, \text{ if } \phi^J = 0 \tag{10.68c}$$
$$\Delta\beta(s_f) = 0$$
$$\Delta\beta(s) \ge -\beta^{(j)}(s), \quad s_{j+1} \le s \le s_f, \tag{10.68d}$$

where

$$R_j := \overline{B}_{g,q_e^{(j)},\beta^{(j)}}^{(j+1)}(s_f) - \overline{C}_{g,q_e^{(j)},\beta^{(j)}}^{(j+1)}(s_f) + \overline{a}_{\phi,q_e^{(j)}}^{(j+1)}(s_f) \tag{10.68e}$$
$$S_j := \overline{C}_{g,q_e^{(j)},\beta^{(j)}}^{(j+1)}(s_f) \tag{10.68f}$$
$$T_j := \frac{1}{2}\overline{C}_{g,q_e^{(j)},\beta^{(j)}}^{(j+1)}(s_{j+1})^T \frac{q_e^{(j)'}(s_{j+1})}{\beta^{(j)}(s_{j+1})} - \overline{E}_{g,q_e^{(j)},\beta^{(j)}}^{(j+1)}(s_{j+1}). \tag{10.68g}$$

The linear optimization problem (10.68a)–(10.68g) can be solved now by the methods developed, e.g., in [13], where for the correction terms $\Delta q_e(s)$, $\Delta\beta(s)$, $s_{j+1} \le s \le s_f$, some box constraints or norm bounds have to be added to (10.68a)–(10.68g). In case of $\Delta\beta(\cdot)$ we may replace, e.g., (10.68d) by the condition

$$-\beta^{(j)}(s) \le \Delta\beta(s) \le \Delta\beta^{\max}, s_{j+1} \le s \le s_f, \tag{10.68d'}$$

with some upper bound $\Delta\beta^{\max}$.

It is easy to see that (10.68a)–(10.68g) can be split up into two separated linear optimization problems for $\Delta q_e(\cdot)$, $\Delta\beta(\cdot)$, respectively. Hence, according to the simple structure of the objective function (10.68a), we observe that

$$\text{sign}\left(\overline{H}_{g,q_e^{(j)},\beta^{(j)}}^{(j+1)}(s)\right), s_{j+1} \le s \le s_f, \text{ and sign}\,(T_j)$$

indicates the points s in the interval $[s_{j+1}, s_f]$ with $\Delta\beta(s) < 0$ or $\Delta\beta(s) > 0$, hence, the points s, $s_{j+1} \le s \le s_f$, where the velocity profile should be decreased/increased. Moreover, using (10.68d'), the optimal correction $\Delta\beta(s)$ is equal to the lower/upper bound in (10.68d') depending on the above-mentioned signs.

Obviously, the correction vectors $\Delta q_e(s)$, $s_{j+1} \le s \le s_f$, for the geometric path in configuration space can be determined in the same way. Similar results are obtained also if we use L_2-norm bounds for the correction terms.

If the pointwise constraints (10.29b), (10.29c) in (10.30a)–(10.30f), (10.30f') are averaged with respect to s, $s_{j+1} \le s \le s_f$, then functions of the type (10.37c) arise, cf. (10.30b'), (10.30c'), which can be linearized again by the same techniques as discussed above. In this case, linear constraints are obtained for $\Delta\beta(s)$, $\Delta q_e(s)$, $s_{j+1} \le s \le s_f$, with constraint functions of the type (10.39a)–(10.39c), cf. also (10.68a).

10.5.3.4 Combination of Discretization and Linearization

Obviously, the methods described briefly in Sects. 10.5.3.1 and 10.5.3.2 can be combined in the following way, cf. Fig. 10.4.

First, by means of discretization (Finite Element Methods), an approximate optimal control law $(\tilde{q}_e^*, \beta^*)$ is searched in a class of finitely generated functions of the type (10.64a), (10.64b). Corresponding to (10.66a), (10.66b), by means of Taylor expansion here the optimal coefficients $\hat{q}_l^*, \hat{\beta}_l^*, l = 1, \ldots, l_q(l_\beta)$, in the corresponding linear combination of type (10.64a), (10.64b) are represented, cf. (10.64c), (10.64d), by

$$\hat{q}_l^*(\zeta, \zeta_f) = \hat{q}_l^*\left(\zeta_0, \zeta_f^{(0)}\right) + \frac{\partial \hat{q}_l^*}{\partial \hat{\zeta}}(\zeta_0, \zeta_f^{(0)})\left(\zeta - \zeta_0, \zeta_f - \zeta_f^{(0)}\right) + \ldots, \tag{10.69a}$$

$$\hat{\beta}_l^*(\zeta, \zeta_f) = \hat{\beta}_l^*\left(\zeta_0, \zeta_f^{(0)}\right) + \frac{\partial \hat{\beta}_l^*}{\partial \hat{\zeta}}\left(\zeta_0, \zeta_f^{(0)}\right)\left(\zeta - \zeta_0, \zeta_f - \zeta_f^{(0)}\right) + \ldots, \tag{10.69b}$$

$l = 1, \ldots, l_q(l_\beta)$. Here, the derivatives

$$\frac{\partial \hat{q}_l^*}{\partial \hat{\zeta}}\left(\zeta_0, \zeta_f^{(0)}\right), \frac{\partial \hat{\beta}_l^*}{\partial \hat{\zeta}}\left(\zeta_0, \zeta_f^{(0)}\right), \ldots, l = 1, \ldots, l_q(l_\beta) \tag{10.69c}$$

can be determined again by sensitivity analysis [48] of a finite dimensional parameter-dependent optimization problem which may be much simpler than the sensitivity analysis of the parameter-dependent variational problem (10.58a)–(10.58f) or (10.59)–(10.59d).

Stating the necessary (and under additional conditions also sufficient) Kuhn-Tucker conditions for the optimal coefficients $\hat{q}_l^*, \hat{\beta}_l^*, l = 1, \ldots, l_q(l_\beta)$, formulas for the derivatives (10.69c) may be obtained by partial differentiation with respect to the complete vector $z = (\zeta, \zeta_f)$ of initial/terminal parameters.

10.6 Online Control Corrections: PD-Controller

We now consider the control of the robot at the j-th stage, i.e., for time $t \geq t_j$, see [1, 2, 4, 7, 14, 15]. In practice we have random variations of the vector p of the model parameters of the robot and its environment, moreover, there are possible deviations of the true initial state $(q_j, \dot{q}_j) := \left(q(t_j), \dot{q}(t_j)\right)$ in configuration space from the corresponding initial values $(\overline{q}_j, \overline{\dot{q}}_j) = \left(\overline{q}_j, q_e^{(j)'}(s_j)\sqrt{\beta_j}\right)$ of the (OSTP) at stage j. Thus, the actual trajectory

$$q(t) = q\Big(t, p_D, q_j, \dot{q}_j, u(\cdot)\Big), t \geq t_j \tag{10.70a}$$

in configuration space of the robot will deviate more or less from the optimal reference trajectory

$$q^{(j)}(t) = q_e^{(j)}\left(s^{(j)}(t)\right) = q\left(t, \overline{p}_D^{(j)}, \overline{q}_j, \overline{\dot{q}}_j, u^{(j)}(\cdot)\right), \tag{10.70b}$$

see (10.45a), (10.45b), (10.46a)–(10.46f), (10.50a), (10.50b) and (10.51c). In the following we assume that the state $\left(q(t), \dot{q}(t)\right)$ in configuration space may be observed for $t > t_j$. Now, in order to define an appropriate control correction (feedback control law), see (10.2) and Fig. 10.3,

$$\Delta u^{(j)}(t) = u(t) - u^{(j)}(t) := \varphi^{(j)}\left(t, \Delta z^{(j)}(t)\right), \quad t \geq t_j, \tag{10.71a}$$

for the compensation of the tracking error

$$\Delta z^{(j)}(t) := z(t) - z^{(j)}(t), \quad z(t) := \begin{pmatrix} q(t) \\ \dot{q}(t) \end{pmatrix}, \quad z^{(j)}(t) := \begin{pmatrix} q^{(j)}(t) \\ \dot{q}^{(j)}(t) \end{pmatrix}, \tag{10.71b}$$

where $\varphi^{(j)} = \varphi^{(j)}(t, \Delta q, \dot{\Delta} q)$ is such a function that

$$\varphi^{(j)}(t, 0, 0) = 0 \text{ for all } t \geq t_j, \tag{10.71c}$$

the trajectories $q(t)$ and $q^{(j)}(t), t \geq t_j$, are embedded into a one-parameter family of trajectories $q = q(t, \epsilon), t \geq t_j, 0 \leq \epsilon \leq 1$, in configuration space which are defined as follows:

Consider first the following initial data for stage j:

$$q_j(\epsilon) := \overline{q}_j + \epsilon \Delta q_j, \quad \Delta q_j := q_j - \overline{q}_j \tag{10.72a}$$

$$\dot{q}_j(\epsilon) := \overline{\dot{q}}_j + \epsilon \dot{\Delta} q_j, \quad \dot{\Delta} q_j := \dot{q}_j - \overline{\dot{q}}_j \tag{10.72b}$$

$$p_D(\epsilon) := \overline{p}_D^{(j)} + \epsilon \Delta p_D, \quad \Delta p_D := p_D - \overline{p}_D^{(j)}, \quad 0 \leq \epsilon \leq 1. \tag{10.72c}$$

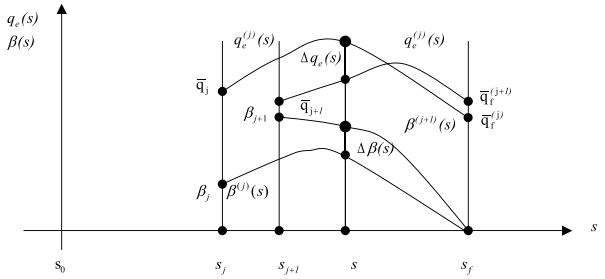


Fig. 10.3 Control of dynamic systems (robots). Here, $z(t) := \left(q(t), \dot{q}(t)\right)$, $z^{(0)}(t) := \left(q^{(0)}(t), \dot{q}^{(0)}(t)\right)$

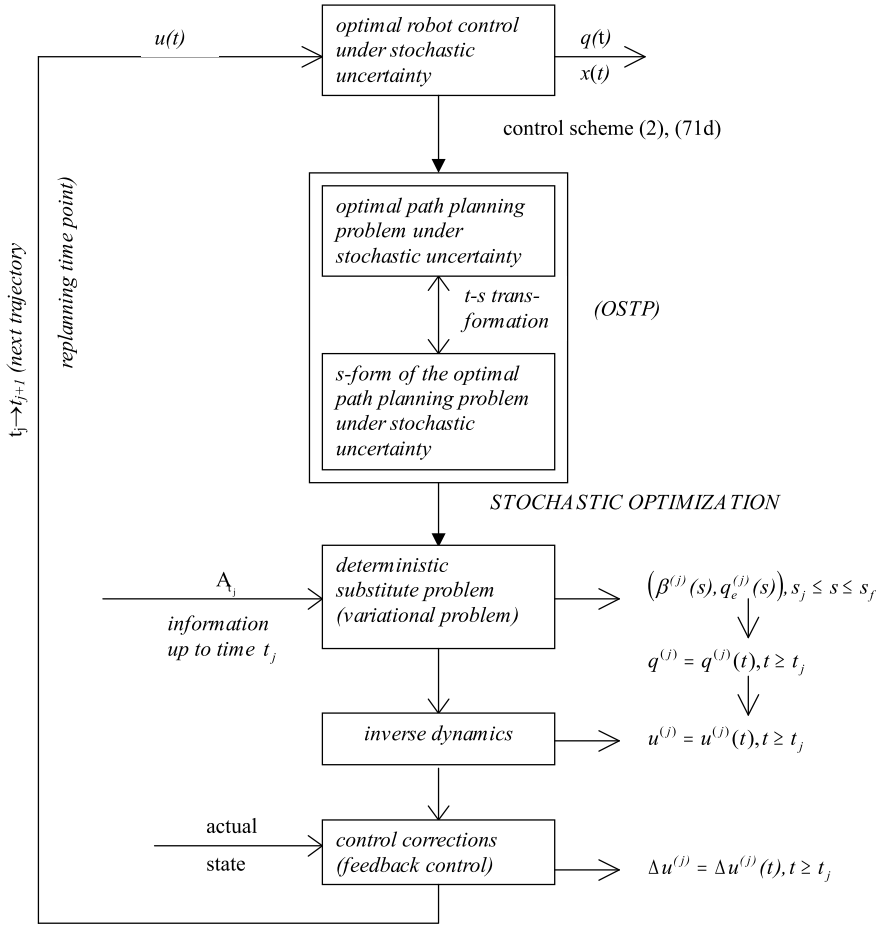


Fig. 10.4 Adaptive optimal stochastic trajectore planning and control (AOSTPC)

Moreover, define the control input $u(t)$, $t \geq t_j$, by (10.71a), hence,

$$\begin{aligned} u(t) &= u^{(j)}(t) + \Delta u^{(j)}(t) \\ &= u^{(j)}(t) + \varphi^{(j)}\left(t, q(t) - q^{(j)}(t), \dot{q}(t) - \dot{q}^{(j)}(t)\right), \ t \geq t_j. \end{aligned} \tag{10.72d}$$

Let then denote

$$q(t, \epsilon) = q\left(t, p_D(\epsilon), q_j(\epsilon), \dot{q}_j(\epsilon), u(\cdot)\right), \quad 0 \leq \epsilon \leq 1, \ t \geq t_j, \tag{10.73}$$

the solution of the following initial value problem consisting of the dynamic equation (10.4a) with the initial values, the vector of dynamic parameters and the total control input $u(t)$ given by (10.72a)–(10.72d) (Fig. 10.4):

$$F\left(p_D(\epsilon), q(t, \epsilon), \dot{q}(t, \epsilon), \ddot{q}(t, \epsilon)\right) = u(t, \epsilon), \quad 0 \leq \epsilon \leq 1, \ t \geq t_j, \tag{10.74a}$$

where

$$q(t_j, \epsilon) = q_j(\epsilon), \quad \dot{q}(t_j, \epsilon) = \dot{q}_j(\epsilon), \tag{10.74b}$$

$$u(t, \epsilon) := u^{(j)}(t) + \varphi^{(j)}\Big(t, q(t, \epsilon) - q^{(j)}(t), \dot{q}(t, \epsilon) - \dot{q}^{(j)}(t)\Big), \tag{10.74c}$$

and $F = F(p_D, q, \dot{q}, \ddot{q})$ is defined, cf. (10.4a), by

$$F(p_D, q, \dot{q}, \ddot{q}) := M(p_D, q)\ddot{q} + h(p_D, q, \dot{q}). \tag{10.74d}$$

In the following we suppose [22] that the initial value problem (10.74a)–(10.74d) has a unique solution $q = q(t, \epsilon)$, $t \geq t_j$, for each parameter value ϵ, $0 \leq \epsilon \leq 1$.

10.6.1 Basic Properties of the Embedding $q(t, \epsilon)$

10.6.1.1 $\epsilon = \epsilon_0 := 0$

Because of condition (10.71c) of the feedback control law $\varphi^{(j)}$ to be determined, and due to the unique solvability assumption of the initial value problem (10.74a)–(10.74d) at the j-th stage, for $\epsilon = 0$ we have that

$$q(t, 0) = q^{(j)}(t), \quad t \geq t_j. \tag{10.75a}$$

$\epsilon = \epsilon_1 := 1$

According to (10.70a), (10.71a)–(10.71c) and (10.72a)–(10.72d),

$$q(t, 1) = q(t) = q\Big(t, p_D, q_j, \dot{q}_j, u(\cdot)\Big), \quad t \geq t_j, \tag{10.75b}$$

is the actual trajectory in configuration space under the total control input $u(t) = u^{(j)}(t) + \Delta u^{(j)}(t)$, $t \geq t_j$, given by (10.72d).

Taylor expansion with respect to ϵ.

Let $\Delta\epsilon = \epsilon_1 - \epsilon_0 = 1$, and suppose that the following property known from parameter-dependent differential equations, cf. [22], holds.

Assumption 10.1 The solution $q = q(t, \epsilon)$, $t \geq t_j$, $0 \leq \epsilon \leq 1$, of the initial value problem (10.72a)–(10.72d) has continuous derivatives with respect to ϵ up to order $\nu > 1$ for all $t_j \leq t \leq t_j + \Delta t_j$, $0 \leq \epsilon \leq 1$, with a certain $\Delta t_j > 0$.

Note that $(t, \epsilon) \to q(t, \epsilon)$, $t \geq t_j$, $0 \leq \epsilon \leq 1$, can be interpreted as a *homotopy* from the reference trajectory $q^{(j)}(t)$ to the actual trajectory $q(t)$, $t \geq t_j$, cf. [39].

Based on the above assumption and (10.75a), (10.75b), by Taylor expansion with respect to ϵ at $\epsilon = \epsilon_0 = 0$, the actual trajectory of the robot can be represented by

$$\begin{aligned} q(t) &= q\Big(t, p_D, q_j, \dot{q}_j, u(\cdot)\Big) = q(t, 1) = q(t, \epsilon_0 + \Delta\epsilon) \\ &= q(t, 0) + \Delta q(t) = q^{(j)}(t) + \Delta q(t), \end{aligned} \tag{10.76a}$$

where the expansion of the tracking error $\Delta q(t)$, $t \geq t_j$, is given by

$$\begin{aligned} \Delta q(t) &= \sum_{l=1}^{\nu-1} \frac{1}{l!} d^l q(t)(\Delta\epsilon)^l + \frac{1}{\nu!} \frac{\partial^\nu q}{\partial \epsilon^\nu}(t, \vartheta)(\Delta\epsilon)^\nu \\ &= \sum_{l=1}^{\nu-1} \frac{1}{l!} d^l q(t) + \frac{1}{\nu!} \frac{\partial^\nu q}{\partial \epsilon^\nu}(t, \vartheta), \quad t \geq t_j. \end{aligned} \tag{10.76b}$$

Here, $\vartheta = \vartheta(t, \nu)$, $0 < \vartheta < 1$, and

$$d^l q(t) := \frac{\partial^l q}{\partial \epsilon^l}(t, 0), \quad t \geq t_j, l = 1, 2, \ldots \tag{10.76c}$$

denote the l-th order differentials of $q = q(t, \epsilon)$ with respect to ϵ at $\epsilon = \epsilon_0 = 0$. Obviously, differential equations for the differentials $d^l q(t)$, $l = 1, 2, \ldots$, may be obtained, cf. [22], by successive differentiation of the initial value problem (10.74a)–(10.74d) with respect to ϵ at $\epsilon_0 = 0$.

Furthermore, based on the Taylor expansion of the tracking error $\Delta q(t)$, $t \geq t_j$, using some stability requirements, the tensorial coefficients $\mathbf{D}_z^l \varphi^{(j)}(t, 0)$, $l = 1, 2, \ldots$, of the Taylor expansion

$$\varphi^{(j)}(t, \Delta z) = \sum_{l=1}^{\infty} \mathbf{D}_z^l \varphi^{(j)}(t, 0) \cdot (\Delta z)^l \tag{10.76d}$$

of the feedback control law $\varphi^{(j)} = \varphi^{(j)}(t, \Delta z)$ can be determined at the same time.

10.6.2 The First-Order Differential dq

Next we have to introduce some definitions. Corresponding to (10.71b) and (10.73) we put

$$z(t,\epsilon) := \begin{pmatrix} q(t,\epsilon) \\ \dot{q}(t,\epsilon) \end{pmatrix}, \quad t \geq t_j,\ 0 \leq \epsilon \leq 1; \tag{10.77a}$$

then, we define the following Jacobians of the function F given by (10.74d):

$$K(p_D, q, \dot{q}, \ddot{q}) := F_q(p_D, q, \dot{q}, \ddot{q}) = \mathbf{D}_q F(p_D, q, \dot{q}, \ddot{q}) \tag{10.77b}$$

$$D(p_D, q, \dot{q}) := F_{\dot{q}}(p_D, q, \dot{q}, \ddot{q}) = h_{\dot{q}}(p_D, q, \dot{q}). \tag{10.77c}$$

Moreover, it is

$$M(p_D, q) = F_{\ddot{q}}(p_D, q, \dot{q}, \ddot{q}), \tag{10.77d}$$

and due to the linear parameterization property of robots, see Remark 10.2, F may be represented by

$$F(p_D, q, \dot{q}, \ddot{q}) = Y(q, \dot{q}, \ddot{q})p_D \tag{10.77e}$$

with a certain matrix function $Y = Y(q, \dot{q}, \ddot{q})$.

By differentiation of (10.74a)–(10.77d) with respect to ϵ, for the partial derivative $\frac{\partial q}{\partial \epsilon}(t,\epsilon)$ of $q = q(t,\epsilon)$ with respect to ϵ we find, cf. (10.71b), the following linear initial value problem **(error differential equation)**

$$\begin{aligned} &Y\Big(q(t,\epsilon), \dot{q}(t,\epsilon), \ddot{q}(t,\epsilon)\Big)\Delta p_D + K\Big(p_D(\epsilon), q(t,\epsilon), \dot{q}(t,\epsilon), \ddot{q}(t,\epsilon)\Big)\frac{\partial q}{\partial \epsilon}(t,\epsilon) \\ &+D\Big(p_D(\epsilon), q(t,\epsilon), \dot{q}(t,\epsilon)\Big)\frac{d}{dt}\frac{\partial q}{\partial \epsilon}(t,\epsilon) + M\Big(p_D(\epsilon), q(t,\epsilon)\Big)\frac{d^2}{dt^2}\frac{\partial q}{\partial \epsilon}(t,\epsilon) \\ &= \frac{\partial u}{\partial \epsilon}(t,\epsilon) = \frac{\partial \varphi^{(j)}}{\partial z}\Big(t, \Delta z^{(j)}(t)\Big)\frac{\partial z}{\partial \epsilon}(t,\epsilon) \end{aligned} \tag{10.78a}$$

with the initial values, see (10.72a), (10.72b),

$$\frac{\partial q}{\partial \epsilon}(t_j,\epsilon) = \Delta q_j, \frac{d}{dt}\frac{\partial q}{\partial \epsilon}(t_j,\epsilon) = \dot{\Delta} q_j. \tag{10.78b}$$

Putting now $\epsilon = \epsilon_0 = 0$, because of (10.71a), (10.71b) and (10.75a), system (10.78a), (10.78b) yields then this system of second-order differential equations for the first-order differential $dq(t) = \frac{\partial q}{\partial \epsilon}(t,0)$:

$$\begin{aligned} &Y^{(j)}(t)\Delta p_D + K^{(j)}(t)dq(t) + D^{(j)}(t)\dot{dq}(t) + M^{(j)}(t)\ddot{dq}(t) \\ &= du(t) = \varphi_z^{(j)}(t,0)dz(t) = \varphi_q^{(j)}(t,0)dq(t) + \varphi_{\dot{q}}^{(j)}(t,0)\dot{dq}(t), t \geq t_j, \end{aligned} \tag{10.79a}$$

with the initial values

$$dq(t_j) = \Delta q_j, \quad \dot{dq}(t_j) = \dot{\Delta} q_j. \tag{10.79b}$$

Here,

$$du(t) := \frac{\partial u}{\partial \epsilon}(t, 0), \tag{10.79c}$$

$$dz(t) := \begin{pmatrix} dq(t) \\ \dot{dq}(t), \end{pmatrix}, \quad \dot{dq} := \frac{d}{dt} dq, \quad \ddot{dq} := \frac{d^2}{dt^2} dq, \tag{10.79d}$$

and the matrices $Y^{(j)}(t)$, $K^{(j)}(t)$, $D^{(j)}(t)$ and $M^{(j)}(t)$ are defined, cf. (10.77b)–(10.77e), by

$$Y^{(j)}(t) := Y\left(q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t)\right) \tag{10.79e}$$

$$K^{(j)}(t) := K\left(\overline{p}_D^{(j)}, q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t)\right) \tag{10.79f}$$

$$D^{(j)}(t) := D\left(\overline{p}_D^{(j)}, q^{(j)}(t), \dot{q}^{(j)}(t)\right), \quad M^{(j)}(t) := M\left(\overline{p}_D^{(j)}, q^{(j)}(t)\right). \tag{10.79g}$$

Local (PD-) control corrections $du = du(t)$ stabilizing system (10.79a), (10.79b) can now be obtained by the following definition of the Jacobian of $\varphi^{(j)}(t, z)$ with respect to z at $z = 0$:

$$\begin{aligned} \varphi_z^{(j)}(t, 0) &:= F_z\left(\overline{p}_D^{(j)}, q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t)\right) - M^{(j)}(t)(K_p, K_d) \\ &= \left(K^{(j)}(t) - M^{(j)}(t) K_p, D^{(j)}(t) - M^{(j)}(t) K_d\right), \end{aligned} \tag{10.80}$$

where $K_p = (\gamma_{pk} \delta_{kv})$, $K_d = (\gamma_{dk} \delta_{kv})$ are positive definite diagonal matrices with positive diagonal elements $\gamma_{pk}, \gamma_{dk} > 0$, $k = 1, \ldots, n$.

Inserting (10.80) into (10.79a), and assuming that $M^{(j)} = M^{(j)}(t)$ is regular [4] for $t \geq t_j$, we find the following linear system of second-order differential equations for $dq = dq(t)$:

$$\ddot{dq}(t) + K_d \dot{dq}(t) + K_p dq(t) = -M^{(j)}(t)^{-1} Y^{(j)}(t) \Delta p_D, \quad t \geq t_j, \tag{10.81a}$$

$$dq(t_j) = \Delta q_j, \quad \dot{dq}(t_j) = \dot{\Delta} q_j. \tag{10.81b}$$

Considering the right-hand side of (10.81a), according to (10.79e), (10.77e) and (10.74d) we have that

$$\begin{aligned} Y^{(j)}(t) \Delta p_D &= Y\left(q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t)\right) \Delta p_D = F\left(\Delta p_D, q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t)\right) \\ &= M\left(\Delta p_D, q^{(j)}(t)\right) \ddot{q}^{(j)}(t) + h\left(\Delta p_D, q^{(j)}(t), \dot{q}^{(j)}(t)\right). \end{aligned} \tag{10.82a}$$

Using the definition (10.50a), (10.50b) of $q^{(j)}(t)$ and the representation (10.19a), (10.19b) of $\dot{q}^{(j)}(t)$, $\ddot{q}^{(j)}(t)$, we get

$$
\begin{aligned}
&Y^{(j)}(t)\Delta p_D \\
&= M\left(\Delta p_D, q_e^{(j)}\left(s^{(j)}(t)\right)\right)\left(q_e^{(j)'}\left(s^{(j)}(t)\right)\frac{1}{2}\beta^{(j)'}\left(s^{(j)}(t)\right)\right. \\
&\left.+q_e^{(j)''}\left(s^{(j)}(t)\right)\beta^{(j)}\left(s^{(j)}(t)\right)\right) \\
&+h\left(\Delta p_D, q_e^{(j)}\left(s^{(j)}(t)\right), q_e^{(j)'}\left(s^{(j)}(t)\right)\sqrt{\beta^{(j)}\left(s^{(j)}(t)\right)}\right). \qquad (10.82b)
\end{aligned}
$$

From (10.20b) we now obtain the following important representations, where we suppose that the feedforward control $u^{(j)}(t)$, $t \geq t_j$, is given by (10.51c), (10.51d).

Lemma 10.1 *The following representations hold:*

(a)

$$Y^{(j)}(t)\Delta p_D = u_e\left(\Delta p_D, s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\right), \quad t \geq t_j; \qquad (10.83a)$$

(b)

$$u^{(j)}(t) = u_e\left(\overline{p}_D^{(j)}, s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\right), \quad t \geq t_j; \qquad (10.83b)$$

(c)

$$u^{(j)}(t) + Y^{(j)}(t)\Delta p_D = u_e\left(p_D, s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\right), \quad t \geq t_j. \qquad (10.83c)$$

Proof The first equation follows from (10.82b) and (10.20b). Equations (10.51c), (10.19a), (10.19b) and (10.20b) yield (10.83b). Finally, (10.83c) follows from (10.83a), (10.83b) and the linear parameterization of robots, cf. Remark 10.2. □

Remark 10.12 Note that according to the transformation (10.20a) of the dynamic equation onto the s-domain, for the control input $u(t)$ we have the representation

$$
\begin{aligned}
u(t) &= u_e\left(p_D, s; q_e(\cdot), \beta(\cdot)\right) \\
&= u_e\left(\overline{p}_D^{(j)}, s; q_e(\cdot), \beta(\cdot)\right) + u_e\left(\Delta p_D, s; q_e(\cdot), \beta(\cdot)\right) \qquad (10.83d)
\end{aligned}
$$

with $s = s(t)$.

Using (10.79d), it is easy to see that (10.81a), (10.81b) can be described also by the first-order initial value problem

$$\dot{dz}(t) = Adz(t) + \begin{pmatrix} 0 \\ \psi^{(j,1)}(t) \end{pmatrix}, \quad t \geq t_j \qquad (10.84a)$$

$$dz(t_j) = \Delta z_j = \begin{pmatrix} q_j - \overline{q}_j \\ \dot{q}_j - \dot{\overline{q}}_j \end{pmatrix}, \qquad (10.84b)$$

where A is the stability or Hurwitz matrix

$$A := \begin{pmatrix} 0 & I \\ -K_p & -K_d \end{pmatrix}, \tag{10.84c}$$

and $\psi^{(j,1)}(t)$ is defined, cf. (10.83a), by

$$\begin{aligned} \psi^{(j,1)}(t) &:= -M^{(j)}(t)^{-1} Y^{(j)}(t) \Delta p_D \\ &= -M^{(j)}(t)^{-1} u_e \left(\Delta p_D, s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot) \right). \end{aligned} \tag{10.84d}$$

Consequently, for the first-order expansion term $dz(t)$ of the deviation $\Delta z^{(j)}(t)$ between the actual state $z(t) = \begin{pmatrix} q(t) \\ \dot{q}(t) \end{pmatrix}$ and the prescribed state $z^{(j)}(t) = \begin{pmatrix} q^{(j)}(t) \\ \dot{q}^{(j)}(t) \end{pmatrix}$, $t \geq t_j$, we have the representation [11, 22]

$$dz(t) = dz^{(j)}(t) = e^{A(t-t_j)} \Delta z_j + \int_{t_j}^{t} e^{A(t-\tau)} \begin{pmatrix} 0 \\ \psi^{(j,1)}(\tau) \end{pmatrix} d\tau. \tag{10.85a}$$

Because of $E\left(\Delta p_D(\omega) | \mathcal{A}_{t_j} \right) = 0$, we have that

$$E\left(\psi^{(j,1)}(t) | \mathcal{A}_{t_j} \right) = 0, \tag{10.85b}$$

$$E\left(dz(t) | \mathcal{A}_{t_j} \right) = e^{A(t-t_j)} E\left(\Delta z_j | \mathcal{A}_{t_j} \right), \quad t \geq t_j, \tag{10.85c}$$

where, see (10.84a), (10.84b),

$$E\left(\Delta z_j | \mathcal{A}_{t_j} \right) = E\left(z(t_j) | \mathcal{A}_{t_j} \right) - \overline{z}_j. \tag{10.85d}$$

It is easy to see that the diagonal elements $\gamma_{dk}, \gamma_{pk} > 0$, $k = 1, \ldots, n$, of the positive definite diagonal matrices K_d, K_p, rep., can be chosen so that the fundamental matrix $\Phi(t, \tau) = e^{A(t-\tau)}$, $t \geq \tau$, is exponentially stable, i.e.,

$$\| \Phi(t, \tau) \| \leq a_0 e^{-\lambda_0 (t-\tau)}, \quad t \geq \tau, \tag{10.86a}$$

with positive constants a_0, λ_0. A sufficient condition for (10.86a) reads

$$\gamma_{dk}, \gamma_{pk} > 0, k = 1, \ldots, n, \text{ and } \gamma_{dk} > 2 \text{ in case of double eigenvalues of } A. \tag{10.86b}$$

Define the generalized variance var $(Z|\mathcal{A}_{t_j})$ of a random vector $Z = Z(\omega)$ given $\mathcal{A}_{t_j}$ by $\text{var}(Z|\mathcal{A}_{t_j}) := E\Big(\|Z - E(Z|\mathcal{A}_{t_j})\|^2|\mathcal{A}_{t_j}\Big)$, and let $\sigma_Z^{(j)} := \sqrt{\text{var}\ (Z|\mathcal{A}_{t_j})}$. Then, for the behavior of the first-order error term $dz(t), t \geq t_j$, we have the following result:

Theorem 10.2 *Suppose that the diagonal matrices K_d, K_p are selected such that (10.86a) holds. Moreover, apply the local (i.e., first order) control correction (PD-controller)*

$$du(t) := \varphi_z^{(j)}(t, 0)dz(t), \tag{10.87a}$$

where $\varphi_z^{(j)}(t, 0)$ is defined by (10.80). Then, the following relations hold:

(a) Asymptotic local stability in the mean:

$$E\Big(dz(t)|\mathcal{A}_{t_j}\Big) \to 0, \quad t \to \infty; \tag{10.87b}$$

(b) Mean absolute first-order tracking error:

$$E\Big(\|dz\||\mathcal{A}_{t_j}\Big) \leq a_0 e^{-\lambda_0(t-t_j)}\sqrt{\sigma_{z(t_j)}^{(j)^2} + \|E\Big(z(t_j)|\mathcal{A}_{t_j}\Big) - \overline{z}_j\|^2}$$
$$+ a_0 \int_{t_j}^{t} e^{-\lambda_0(t-\tau)}\sqrt{E\left(\|\psi^{(j,1)}(\tau)\|^2|\mathcal{A}_{t_j}\right)}\, d\tau, \quad t \geq t_j, \tag{10.87c}$$

where

$$E\left(\|\psi^{(j,1)}(t)\|^2|\mathcal{A}_{t_j}\right) \leq \|M^{(j)}(t)^{-1}\|^2\sigma_{u_e}^{(j)2}\Big(s^{(j)}(t)\Big), \tag{10.87d}$$
$$\sigma_{u_e}^{(j)^2}\Big(s^{(j)}(t)\Big) \leq \|Y^{(j)}(t)\|^2\ \text{var}\ \Big(p_D(\cdot)|\mathcal{A}_{t_j}\Big) \tag{10.87e}$$

with

$$\sigma_{u_e}^{(j)}(s) := \sqrt{\text{var}\ \Big(u_e\Big(p_D(\cdot), s; q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\Big)|\mathcal{A}_{t_j}\Big)}, \quad s_j \leq s \leq s_f. \tag{10.87f}$$

Proof Follows from (10.85a), (10.83a)–(10.83d) and the fact that by Jensen's inequality $E\sqrt{X(\omega)} \leq \sqrt{EX(\omega)}$ for a nonnegative random variable $X = X(\omega)$. □

Note that $\sigma_{u_e}^{(j)2}\Big(s^{(j)}(t)\Big)$ can be interpreted as the risk of the feedforward control $u^{(j)}(t),\ t \geq t_j$. Using (10.70b), (10.79g), (10.87d), (10.87e) and then changing variables $\tau \to s$ in the integral in (10.87c), we obtain the following result:

Theorem 10.3 *Let denote $t^{(j)} = t^{(j)}(s)$, $s \geq s_j$, the inverse of the parameter transformation $s^{(j)} = s^{(j)}(t)$, $t \geq t_j$. Under the assumptions of Theorem 10.2, the following inequality holds for $t_j \leq t \leq t_f^{(j)}$:*

$$E\Big(\|dz(t)\||\mathcal{A}_{t_j}\Big) \leq a_0 e^{-\lambda_0(t-t_j)} \sqrt{\sigma_{z(t_j)}^{(j)^2} + \|E\big(z(t_j)|\mathcal{A}_{t_j}\big) - \overline{z}_j\|^2}$$
$$+ \int_{s_j}^{s^{(j)}(t)} \frac{a_0 e^{-\lambda_0\left(t-t^{(j)}(s)\right)} \|M\left(\overline{p}_D^{(j)}, q_e^{(j)}(s)\right)^{-1}\|}{\sqrt{\beta^{(j)}(s)}}\, \sigma_{u_e}^{(j)}(s)\, ds. \tag{10.88a}$$

The minimality or boundedness of the right-hand side of (10.88a), hence, the robustness [12] of the present control scheme, is shown next:

Corollary 10.1 *The meaning of the above inequality (10.88a) follows from the following important minimality/boundedness properties depending on the chosen substitute problem in (OSTP) for the trajectory planning problem under stochastic uncertainty:*

(i) *The error contribution of the initial value $\overline{z}_j$ takes a minimum for $\overline{z}_j := E\Big(z(t_j)|\mathcal{A}_{t_j}\Big)$, cf. (10.45a), (10.45b).*

(ii) *The factor λ_0 can be increased by an appropriate selection of the matrices K_p, K_d;*

(iii)

$$\underline{c}_M \leq \|M\left(p_D^{(j)}, q_e^{(j)}(s)\right)^{-1}\| \leq \overline{c}_M, \quad s_j \leq s \leq s_f, \tag{10.88b}$$

with positive constants $\underline{c}_M, \overline{c}_M > 0$. This follows from the fact that the mass matrix is always positive definite [4].

(iv)

$$\int_{s_j}^{s^{(j)}(t)} \frac{ds}{\sqrt{\beta^{(j)}(s)}} \leq \int_{s_j}^{s_f} \frac{ds}{\sqrt{\beta^{(j)}(s)}} = t_f^{(j)} - t_j, \tag{10.88c}$$

where according to (OSTP), for minimum-time and related substitute problems, the right-hand side is a minimum.

(v) *Depending on the chosen substitute problem in (OSTP), the generalized variance $\sigma_{u_e}^{(j)}(s)$, $s_f \leq s \leq s_f$, is bounded pointwise by an appropriate upper risk level, or $\sigma_{u_e}^{(j)}(\cdot)$ minimized in a certain weighted mean sense.*

For the minimality or boundedness of the generalized variance $\sigma_{u_e}^{(j)^2}(s)$, $s_j \leq s \leq s_f$, mentioned above, we give the following examples:

Working with the probabilistic control constraints (10.31a) and assuming that the vectors u^c and ρ_u are fixed, see (10.31d), according to (10.31f) we find that (10.31a) can be guaranteed by

$$\sigma_{u_e}^{(j)^2}(s) + \|\overline{u}_e^{(j)}(s) - u^c\|^2 \le (1-\alpha_u) \min_{1\le k\le n} \rho_{uk}^2, \quad s_j \le s \le s_f, \tag{10.88d}$$

where $\overline{u}_e^{(j)}(s) := u_e\Big(\overline{p}_D^{(j)}, s; q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\Big)$. Hence, with (10.88d) we have then the condition

$$\sigma_{u_e}^{(j)^2}(s) \le (1-\alpha_u) \min_{1\le k\le n} \rho_{u_e}^2, \ s_j \le s \le s_f, \tag{10.88d'}$$

cf. (10.88a). Under special distribution assumptions for $p_D(\omega)$ more exact explicit deterministic conditions for (10.31a) may be derived, see Remark 10.2.

If minimum force and moment should be achieved along the trajectory, hence, if $\phi = 0$ and $L = \|u(t)\|^2$, see (10.6), then, according to substitute problem (10.30a)–(10.30f), (10.30f') we have the following minimality property:

$$\int_{s_j}^{s_f} \left(\sigma_{u_e}^{(j)^2}(s) + \left\|\overline{u}_e^{(j)}(s)\right\|^2\right) \frac{ds}{\sqrt{\beta^{(j)}(s)}} = \min_{q_e(\cdot),\beta(\cdot)} E\Big(\int_{t_j}^{t_f} \|u(t)\|^2\, dt | \mathcal{A}_{t_j}\Big). \tag{10.88e}$$

Mean/variance condition for u_e: Condition (10.30b) in substitute problem (10.30a)–(10.30f), (10.30f') may read in case of fixed bounds $u^{\min}, u^{\max}$ for the control $u(t)$ as follows:

$$u^{\min} \le u_e\Big(\overline{p}_D^{(j)}, s; q_e(\cdot), \beta(\cdot)\Big) \le u^{\max}, \quad s_j \le s \le s_f \tag{10.88f}$$

$$\sigma_{u_e}^{(j)}(s) \le \sigma_{u_e}^{\max}, \quad s_f \le s \le s_f \tag{10.88g}$$

with a given upper bound $\sigma_{u_e}^{\max}$, cf. (10.88d').

According to Theorem 10.3, further stability results, especially the convergence

$$E\Big(\|dz(t)\| | \mathcal{A}_{t_j}\Big) \to 0 \quad \text{for } j \to \infty, t \to \infty \tag{10.89a}$$

of the mean absolute first-order tracking error can be obtained if, by using a suitable update law [1, 2, 4, 11] for the parameter estimates, hence, for the a posteriori distribution $P(\cdot|\mathcal{A}_{t_j})$, we have that, see (10.87f),

$$\text{var}\left(p_D(\cdot)|\mathcal{A}_{t_j}\right) = \text{var}\left(p_D(\cdot)|\mathcal{A}_{t_j}\right) \to 0 \quad \text{for } j \to \infty. \tag{10.89b}$$

10.6.3 The Second-Order Differential d^2q

In order to derive a representation of the second-order differential d^2q, Eq. (10.78a) for $\dfrac{\partial q}{\partial \epsilon}(t,\epsilon)$ is represented as follows:

$$F_{p_D}\Delta p_D + F_z\frac{\partial z}{\partial \epsilon} + F_{\ddot{q}}\frac{\partial \ddot{q}}{\partial \epsilon} = \frac{\partial u}{\partial \epsilon} = \varphi_z^{(j)}\frac{\partial z}{\partial \epsilon}, \tag{10.90a}$$

where $F = F(p_D, z, \ddot{q})$, $z = \binom{q}{\dot{q}}$, is given by (10.74d), see also (10.77e), and therefore

$$F_{p_D} = F_{p_D}(q,\dot{q},\ddot{q}) = Y(q,\dot{q},\ddot{q}), \quad F_{\ddot{q}} = F_{\ddot{q}}(p_D,q) = M(p_D,q) \tag{10.90b}$$

$$F_z = F_z(p_D,q,\dot{q},\ddot{q}) = \big(F_q, F_{\dot{q}}\big) = \Big(K(p_D,q,\dot{q},\ddot{q}), D(p_D,q,\dot{q})\Big). \tag{10.90c}$$

Moreover, we have that

$$\varphi^{(j)} = \varphi^{(j)}\Big(t, z - z^{(j)}(t)\Big), \quad z = \begin{pmatrix} q(t,\epsilon) \\ \dot{q}(t,\epsilon)\end{pmatrix}, \quad p_D = p_D(\epsilon). \tag{10.90d}$$

By differentiation of (10.90a) with respect to ϵ, we obtain

$$\begin{aligned} &2F_{p_Dz}\cdot\left(\Delta p_D, \frac{\partial z}{\partial \epsilon}\right) + 2F_{p_D\ddot{q}}\cdot\left(\Delta p_D, \frac{\partial \ddot{q}}{\partial \epsilon}\right) + F_{zz}\cdot\left(\frac{\partial z}{\partial \epsilon}, \frac{\partial z}{\partial \epsilon}\right) \\ &+2F_{z\ddot{q}}\cdot\left(\frac{\partial z}{\partial \epsilon}, \frac{\partial \ddot{q}}{\partial \epsilon}\right) + F_z\frac{\partial^2 z}{\partial \epsilon^2} + F_{\ddot{q}}\frac{\partial^2 \ddot{q}}{\partial \epsilon^2} \\ &\qquad = \varphi_{zz}^{(j)}\cdot\left(\frac{\partial z}{\partial \epsilon}, \frac{\partial z}{\partial \epsilon}\right) + \varphi_z^{(j)}\frac{\partial^2 z}{\partial \epsilon^2}, \end{aligned} \tag{10.91a}$$

with the second-order partial derivatives

$$F_{p_Dz} = F_{p_Dz}(z,\ddot{q}), \qquad F_{p_D\ddot{q}} = F_{p_D\ddot{q}}(q) = M_{p_D}(q) \tag{10.91b}$$

$$F_{zz} = F_{zz}(p_D,z,\ddot{q}), \qquad F_{z\ddot{q}} = F_{z\ddot{q}}(p_D,z) = \Big(M_q(p_D,q), 0\Big). \tag{10.91c}$$

Moreover, differentiation of (10.78b) with respect to ϵ yields the initial values

$$\frac{\partial^2 q}{\partial \epsilon^2}(t_j,\epsilon) = 0, \quad \frac{d}{dt}\frac{\partial^2 q}{\partial \epsilon^2}(t_j,\epsilon) = \frac{\partial^2 \dot{q}}{\partial \epsilon^2}(t_j,\epsilon) = 0 \tag{10.91d}$$

for $\dfrac{\partial^2 z}{\partial \epsilon^2} = \dfrac{\partial^2 z}{\partial \epsilon^2}(t,\epsilon) = \left(\dfrac{\partial^2 q}{\partial \epsilon^2}(t,\epsilon), \dfrac{\partial^2 \dot{q}}{\partial \epsilon^2}(t,\epsilon)\right),\ t \geq t_j,\ 0 \leq \epsilon \leq 1.$

Putting now $\epsilon = 0$, from (10.91a) we obtain the following differential equation for the second-order differential $d^2q(t) = \frac{\partial^2 q}{\partial \epsilon^2}(t, 0)$ of $q = q(t, \epsilon)$:

$$\begin{aligned}
&K^{(j)}(t)d^2q(t) + D^{(j)}(t)\frac{d}{dt}d^2q(t) + M^{(j)}(t)\frac{d^2}{dt^2}d^2q(t)\\
&+\left(F_{zz}\left(\overline{p}_D^{(j)}, q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t)\right) - \varphi_{zz}^{(j)}(t, 0)\right)\cdot\left(dz(t), dz(t)\right)\\
&-\varphi_z^{(j)}(t, 0)d^2z(t) = -2F_{p_Dz}^{(j)}(t)\cdot\left(\Delta p_D, dz(t)\right)\\
&+2F_{p_D\ddot{q}}^{(j)}(t)\cdot\left(\Delta p_D, \ddot{d}q(t)\right) + 2F_{z\ddot{q}}^{(j)}(t)\cdot\left(dz(t), \ddot{d}q(t)\right).
\end{aligned} \tag{10.92a}$$

Here, we set

$$d^2z(t) := \begin{pmatrix} d^2q(t) \\ \frac{d}{dt}d^2q(t) \end{pmatrix}, \tag{10.92b}$$

and the vectorial Hessians $F_{p_Dz}^{(j)}(t)$, $F_{p_D\ddot{q}}^{(j)}(t)$, $F_{z\ddot{q}}^{(j)}(t)$ follow from (10.91b) by inserting there the argument $(p_D, q, \dot{q}, \ddot{q}) := \left(\overline{p}_D^{(j)}, q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t)\right)$. Furthermore, (10.91d) yields the following initial condition for $d^2q(t)$

$$d^2q(t_j) = 0, \quad \frac{d}{dt}d^2q(t_j) = 0. \tag{10.92c}$$

According to (10.92a) we define now, cf. (10.80), the second-order derivative of $\varphi^{(j)}$ with respect to z at $\Delta z = 0$ by

$$\varphi_{zz}^{(j)}(t, 0) := F_{zz}\left(\overline{p}_D^{(j)}, q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t)\right), \quad t \geq t_j. \tag{10.93}$$

Using the definition (10.80) of the Jacobian of $\varphi^{(j)}$ with respect to z at $\Delta z = 0$, for $d^2q = d^2q(t)$, $t \geq t_j$, we find the following the initial value problem

$$\begin{aligned}
&\frac{d^2}{dt^2}d^2q(t) + K_d\frac{d}{dt}d^2q(t) + K_pd^2q(t)\\
&= -M^{(j)}(t)^{-1}\widetilde{\mathbf{D}^2F}^{(j)}(t)\cdot\left(\Delta p_D, dz(t), \ddot{d}q(t)\right)^2, \quad t \geq t_j,
\end{aligned} \tag{10.94a}$$

$$d^2q(t_j) = 0, \frac{d}{dt}d^2q(t_j) = 0, \tag{10.94b}$$

where the sub-Hessian

$$\widetilde{\mathbf{D}^2F}^{(j)}(t) := \widetilde{\mathbf{D}^2F}\left(\overline{p}_D^{(j)}, q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t)\right), \tag{10.94c}$$

of F results from the Hessian of F by replacing the diagonal block F_{zz} by zero. Of course, we have that, cf. (10.92a),

$$\widetilde{\nabla^2 F}^{(j)} \cdot \left(\Delta_{p_D}, dz(t), \ddot{d}q(t)\right)^2 = 2\left(F^{(j)}_{p_D z}(t) \cdot \left(\Delta p_D, dz(t)\right)\right.$$
$$\left. + F^{(j)}_{p_D \ddot{q}}(t) \cdot \left(\Delta p_D, \ddot{d}q(t)\right) + F^{(j)}_{z\ddot{q}}(t) \cdot \left(\ddot{d}q(t), dz(t)\right)\right). \quad (10.94d)$$

Comparing now the initial value problems (10.81a), (10.81b) and (10.94a), (10.94b) for the first and second-order differential of $q = q(t, \epsilon)$, we recognize that the linear second-order differential equations have—up to the right-hand side—exactly the same form.

According to (10.84a)–(10.84d) and (10.85a) we know that the first-order expansion terms

$$\left(dz(t), \ddot{d}q(t)\right) = \left(dq(t), \dot{d}q(t), \ddot{d}q(t)\right), \quad t \geq t_j,$$

in the tracking error depend linearly on the error term

$$\Delta\theta^{(j)} = \Delta\theta^{(j)}(t) := \begin{pmatrix} e^{A(t-t_j)}\Delta z_j \\ \Delta p_D \end{pmatrix} \left(\to \begin{pmatrix} 0 \\ \Delta p_D \end{pmatrix}, t \to \infty\right) \quad (10.95)$$

corresponding to the variations/disturbances of the initial values $(q_j, \dot{q}_j)$ and dynamic parameters p_D. Consequently, we have this observation:

Lemma 10.2 *The right-hand side*

$$\psi^{(j,2)}(t) := -M^{(j)}(t)^{-1}\widetilde{\mathbf{D}^2 F}^{(j)}(t) \cdot \left(\Delta p_D, dz(t), \ddot{d}q(t)\right)^2, t \geq t_j, \quad (10.96)$$

of the error differential equation (10.94a) for $d^2q(t)$ is quadratic in the error term $\Delta\theta^{(j)}(\cdot)$.

According to (10.94a)–(10.94c), the second-order expansion term

$$d^2 z(t) = \left(d^2 q(t), \frac{d}{dt} d^2 q(t)\right), \quad t \geq t_j,$$

of the Taylor expansion of the tracking error can be represented again by the solution of the system of linear differential equations (10.84a)–(10.84c), where now $\psi^{(j,1)}(t)$ is replaced by $\psi^{(j,2)}(t)$ defined by (10.96), and the initial values are given by $d^2 z(t_j) = 0$. Thus, applying again solution formula (10.85a), we find

$$d^2 z(t) = \int_{t_j}^{t} e^{A(t-\tau)} \begin{pmatrix} 0 \\ \psi^{(j,2)}(\tau) \end{pmatrix} d\tau. \quad (10.97)$$

From (10.95)–(10.97) and Lemma 10.2 we get now the following result.

Theorem 10.4 *The second-order tracking error expansion terms* $\left(d^2 z(t), \frac{d^2}{dt^2} d^2 q(t)\right) = \left(d^2 q(t), \frac{d}{dt} d^2 q(t), \frac{d^2}{dt^2} d^2 q(t)\right), t \geq t_j$, *depend*

(i) quadratically on the first-order error terms $\left(\Delta p_D, dz(t), \ddot{d}q(t)\right)$ *and*

(ii) quadratically on the error term $\Delta\theta^{(j)}(\cdot)$ *corresponding to the variations/ disturbances of the initial values and dynamics parameters.*

Because of (10.97), the stability properties of the second-order tracking error expansion term $d^2q(t), t \geq t_j$, are determined again by the matrix exponential function $\Phi(t,\tau) = e^{A(t-\tau)}$ and the remainder $\psi^{(j,2)}(t)$ given by (10.96).

According to Theorem 10.3 and Corollary 10.1 we know that the disturbance term $\psi^{(j,1)}(t)$ of (10.81a), (10.84a) is reduced directly by control constraints (for u_e) present in (OSTP). Concerning the next disturbance term $\psi^{(j,2)}(t)$ of (10.94a), by (10.96) we note first that a reduction of the 1st order error terms $\left(\Delta p_D, dz(\cdot), \ddot{d}q(\cdot)\right)$ yields a reduction of $\psi^{(j,2)}$ and by (10.97) also a reduction of $d^2z(t)$, $\dfrac{d^2}{dt^2}d^2q(t), t \geq t_j$. Comparing then definitions (10.84d) and (10.96) of the disturbances $\psi^{(j,1)}$, $\psi^{(j,2)}$, we observe that, corresponding to $\psi^{(j,1)}$, certain terms in $\psi^{(j,2)}$ depend only on the reference trajectory $q^{(j)}(t), t \geq t_j$, of stage j. Hence, this observation yields the following result.

Theorem 10.5 *The disturbance* $\psi^{(j,2)}$ *of (10.94a), and consequently also the second-order tracking error expansion terms* $d^2q(t)$, $\dfrac{d}{dt}d^2q(t)$, $\dfrac{d^2}{dt^2}d^2q(t)$, $t \geq t_j$, *can be diminished by*

(i) reducing the first-order error terms $\left(\Delta p_D, dz(\cdot), \ddot{q}(\cdot)\right)$, *and by*

(ii) taking into (OSTP) additional conditions for the unknown functions $q_e(s)$, $\beta(s)$, $s_j \leq s \leq s_f$, *guaranteeing that (the norm of) the sub-Hessian* $\widetilde{\mathbf{D}^2F}^{(j)}(t)$, $t \geq t_j$, *fulfills a certain minimality or boundedness condition.*

Proof Follows from definition (10.96) of $\psi^{(j,2)}$ and representation (10.97) of the second-order tracking error expansion term d^2z. □

10.6.4 Third and Higher Order Differentials

By further differentiation of equations (10.91a), (10.91d) with respect to ϵ and by putting $\epsilon = 0$, also the third and higher order differentials $d^lq(t)$, $t \geq t_j, l \geq 3$, can be obtained. We observe that the basic structure of the differential equations for the differentials $d^lq, l \geq 1$, remains the same. Hence, by induction for the differentials $d^lz, l \geq 1$, we have the following representation:

Theorem 10.6 *Defining the tensorial coefficients of the Taylor expansion (10.76d) for the feedback control law* $\varphi^{(j)} = \varphi^{(j)}(t, \Delta z)$ *by*

$$\mathbf{D}_z^l\varphi^{(j)}(t,0) := \mathbf{D}_z^l F\left(\overline{p}_D^{(j)}, q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t)\right), \quad t \geq t_j, \quad l = 1,2,\ldots, \tag{10.98}$$

the differentials $d^l z(t) = \left(d^l q(t), \dfrac{d}{dt} d^l q(t) \right)$, $t \geq t_j$, *may be represented by the systems of linear differential equations*

$$\frac{d}{dt} d^l z(t) = A d^l z(t) + \begin{pmatrix} 0 \\ \psi^{(j,l)}(t) \end{pmatrix}, \quad t \geq t_j, l = 1, 2, \ldots, \tag{10.99a}$$

with the same system matrix A and the disturbance terms $\psi^{(j,l)}(t)$, $t \geq t_j$, *given by*

$$\psi^{(j,l)}(t) = -M^{(j)}(t)^{-1} \pi \Big(\widetilde{\mathbf{D}^{\lambda} F}^{(j)}(t), 2 \leq \lambda \leq l; \Delta p_D, d^j z(t), \frac{d^2}{dt^2} d^k q(t), 1 \leq j, k \leq l-1 \Big), \; l \geq 2, \tag{10.99b}$$

where π *is a polynomial in the variables* Δp_D *and* $d^j z(t)$, $\dfrac{d^2}{dt^2} d^k q(t)$, $1 \leq j, k \leq l-1$, *having coefficients from the sub-operators* $\widetilde{\mathbf{D}^{\lambda} F}^{(j)}(t)$ *of* $\mathbf{D}^{\lambda} F^{(j)}(t)$ *containing mixed partial derivatives of F with respect to* $\Delta p_D, z, \ddot{q}$ *of order* $\lambda = 2, 3, \ldots, l-1$ *at* $(p_D, q, \dot{q}, \ddot{q}) = \left(\overline{p_D}^{(j)}, q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t) \right)$ *such that the disturbance* $\psi^{(j,l)}$ *is a polynomial of order l with respect to the error term* $\Delta\theta^{(j)}(\cdot)$.

According to (10.74d), (10.4a)–(10.4d) and Remark 10.2 we know that the vector function $F = F(p_D, q, \dot{q}, \ddot{q})$ is linear in p_D, linear in $\ddot{q}$ and quadratic in $\dot{q}$ (supposing case 10.4c)) and analytical with respect to q. Hence, working with a polynomial approximation with respect to q, we may assume that

$$\mathbf{D}^l F^{(j)}(t) = \mathbf{D}^l F \left(\overline{p_D}^{(j)}, q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t) \right) \approx 0, \quad t \geq t_j, l \geq l_0 \tag{10.100}$$

for some index l_0.

According to the expansion (10.76d) of the feedback control law $\varphi^{(j)} = \varphi^{(j)}(t, \Delta z)$, definition (10.98) of the corresponding coefficients and Theorem 10.5 we have now this robustness [12] result.

Theorem 10.7 *The Taylor expansion (10.76d) of the feedback control law* $\varphi^{(j)} = \varphi^{(j)}(t, \Delta z)$ *stops after a finite number* $(\leq l_0)$ *of terms. Besides the conditions for* u_e *contained automatically in (OSTP) via the control constraints, the mean absolute tracking error* $E\Big(\|\Delta z(t)\| \, | \mathcal{A}_{t_j} \Big)$ *can be diminished further by including additional conditions for the functions* $\Big(q_e(s), \beta(s) \Big)$, $s_j \leq s \leq s_f$, *in (OSTP) which guarantee a minimality or boundedness condition for (the norm of) the sub-operators of mixed partial derivatives* $\widetilde{\mathbf{D}^2 F}^{(j)}(t)$, $t \geq t_j, \lambda = 2, 3, \ldots, l_0$.

10.7 Online Control Corrections: PID Controllers

Corresponding to Sect. 10.5, Eqs. (10.71a)–(10.71c), at stage j we consider here control corrections, hence, feedback control laws, of the type

$$\Delta u^{(j)}(t) := u(t) - u^{(j)}(t) = \varphi^{(j)}\big(t, \Delta z^{(j)}(t)\big), \quad t \geq t_j, \tag{10.101a}$$

where

$$\Delta z^{(j)}(t) := z(t) - z^{(j)}(t), \quad t \geq t_j,$$

is the tracking error related now to the state vector

$$z(t) := \begin{pmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{pmatrix} = \begin{pmatrix} q(t) \\ \int\limits_{t_j}^{t} q(s)ds \\ \dot{q}(t) \end{pmatrix}, \quad t \geq t_j \tag{10.101b}$$

$$z^{(j)}(t) := \begin{pmatrix} z_1^{(j)}(t) \\ z_2^{(j)}(t) \\ z_3^{(j)}(t) \end{pmatrix} = \begin{pmatrix} q^{(j)}(t) \\ \int\limits_{t_j}^{t} q^{(j)}(s)ds \\ \dot{q}^{(j)}(t) \end{pmatrix}, \quad t \geq t_j. \tag{10.101c}$$

Furthermore,

$$\begin{aligned} \varphi^{(j)} &= \varphi^{(j)}\big(t, \Delta z(t)\big) \\ &= \varphi^{(j)}\big(t, \Delta z_1(t), \Delta z_2(t), \Delta z_3(t)\big), \quad t \geq t_j, \end{aligned} \tag{10.101d}$$

is a feedback control law such that

$$\varphi^{(j)}(t, 0, 0, 0) = 0, \quad t \geq t_j. \tag{10.101e}$$

Of course, we have

$$\Delta z(t) = \Delta z^{(j)}(t) = \begin{pmatrix} q(t) - q^{(j)}(t) \\ \int\limits_{t_j}^{t} \big(q(s) - q^{(j)}(s)\big)ds \\ \dot{q}(t) - \dot{q}^{(j)}(t) \end{pmatrix}, \quad t \geq t_j. \tag{10.102}$$

Corresponding to Sect. 10.5, the trajectories $q = q(t)$ and $q^{(j)} = q^{(j)}(t)$, $t \geq t_j$, are embedded into a one-parameter family of trajectories $q = q(t, \varepsilon)$, $t \geq t_j$, $0 \leq \varepsilon \leq 1$, in configuration space which is defined as follows.

At stage j, here we have the following **initial data**:

$$q_j(\varepsilon) := \overline{q}_j + \varepsilon \Delta q_j, \qquad \Delta q_j := q_j - \overline{q}_j \tag{10.103a}$$

$$\dot{q}_j(\varepsilon) := \dot{\overline{q}}_j + \varepsilon \Delta \dot{q}_j, \qquad \Delta \dot{q}_j := \dot{q}_j - \dot{\overline{q}}_j \tag{10.103b}$$

$$p_D(\varepsilon) := \overline{p}_D^{(j)} + \varepsilon \Delta p_D, \qquad \Delta p_D := p_D - \overline{p}_D^{(j)}, \tag{10.103c}$$

$0 \le \varepsilon \le 1$. Moreover, the control input $u = u(t)$, $t \ge t_j$, is defined by

$$\begin{aligned} u(t) &= u^{(j)}(t) + \Delta u^{(j)}(t) \\ &= u^{(j)}(t) + \varphi^{(j)}\Big(t, q(t) - q^{(j)}(t), \int_{t_j}^{t} \big(q(s) - q^{(j)}(s)\big) ds, \\ &\quad \dot{q}(t) - \dot{q}^{(j)}(t)\Big), \quad t \ge t_j. \end{aligned} \tag{10.103d}$$

Let denote now

$$q(t, \varepsilon) = q\big(t, p_D(\varepsilon), q_j(\varepsilon), \dot{q}_j(\varepsilon), u(\cdot)\big), \quad t \ge t_j, \ 0 \le \varepsilon \le 1, \tag{10.104}$$

the solution of the following initial value problem based on the dynamic equation (10.4a) having the initial values and total control input $u(t)$ given by (10.103a)–(10.103d):

$$F\big(p_D(\varepsilon), q(t, \varepsilon), \dot{q}(t, \varepsilon), \ddot{q}(t, \varepsilon)\big) = u(t, \varepsilon), \quad t \ge t_j, 0 \le \varepsilon \le 1, \tag{10.105a}$$

where

$$q(t_j, \varepsilon) := q_j(\varepsilon), \dot{q}(t_j, \varepsilon) = \dot{q}_j(\varepsilon) \tag{10.105b}$$

$$\begin{aligned} u(t, \varepsilon) &:= u^{(j)}(t) + \varphi^{(j)}\Big(t, q(t, \varepsilon) - q^{(j)}(t), \int_{t_j}^{t} \big(q(s, \varepsilon) - q^{(j)}(s)\big) ds, \\ &\quad \dot{q}(t, \varepsilon) - \dot{q}^{(j)}(t)\Big), \end{aligned} \tag{10.105c}$$

and the vector function $F = F(p_D, q, \dot{q}, \ddot{q})$ is again defined by

$$F(p_D, q, \dot{q}, \ddot{q}) = M(p_D, q)\ddot{q} + h(p_D, q, \dot{q}), \tag{10.105d}$$

cf. (10.105a).

In the following we assume that problem (10.105a)–(10.105d) has a unique solution $q = q(t, \varepsilon)$, $t \ge t_j$, for each parameter ε, $0 \le \varepsilon \le 1$.

10.7.1 Basic Properties of the Embedding $q(t, \varepsilon)$

According to (10.103a)–(10.103c), for $\varepsilon = 0$ system (10.105a)–(10.105d) reads

$$F\big(\overline{p}_D^{(j)}, q(t,0), \dot{q}(t,0), \ddot{q}(t,0)\big) = u^{(j)}(t) + \varphi^{(j)}\Big(t, q(t,0) - q^{(j)}(t), \int_{t_j}^{t} \big(q(s,0) - q^{(j)}(s)\big)ds, \dot{q}(t,0) - \dot{q}^{(j)}(t)\Big), \quad t \geq t_j, \tag{10.106a}$$

where

$$q(t_j, 0) = \overline{q}_j, \quad \dot{q}(t_j, 0) = \overline{\dot{q}}_j. \tag{10.106b}$$

Since, due to (OSTP),

$$q^{(j)}(t_j) = \overline{q}_j, \quad \dot{q}^{(j)}(t_j) = \overline{\dot{q}_j}$$

and

$$F\big(\overline{p}_D^{(j)}, q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t)\big) = u^{(j)}(t), \quad t \geq t_j,$$

according to the unique solvability assumption for (10.105a)–(10.105d) and condition (10.101e) we have

$$q(t, 0) = q^{(j)}(t), \quad t \geq t_j. \tag{10.107}$$

For $\varepsilon = 1$, from (10.103a)–(10.103c) and (10.105a)–(10.105d) we obtain the system

$$F\big(p_D, q(t,1), \dot{q}(t,1), \ddot{q}(t,1)\big) = u^{(j)}(t) + \varphi^{(j)}\Big(t, q(t,1) - q^{(j)}(t), \int_{t_j}^{t} \big(q(s,1) - q^{(j)}(s)\big)ds, \dot{q}(t,1) - \dot{q}^{(j)}(t)\Big), \quad t \geq t_1 \tag{10.108a}$$

with

$$q(t_j, 1) = q_j, \quad \dot{q}(t_j, 1) = \dot{q}_j. \tag{10.108b}$$

However, since the control input is defined by (10.103d), again due to the unique solvability property of (10.105a)–(10.105d), (10.108a), (10.108b) yields

$$q(t, 1) = q(t), \quad t \geq t_j, \tag{10.109}$$

where $q = q(t)$ denotes the **actual trajectory.**

Remark 10.13 The integro-differential equation (10.105a)–(10.105d) can be easily converted into an ordinary initial value problem. Indeed, using the state variables

$$z(t,\varepsilon) = \begin{pmatrix} q(t;\varepsilon) \\ q_I(t,\varepsilon) \\ \dot{q}(t,\varepsilon) \end{pmatrix}, \quad t \geq t_j,$$

where $q_I = q_I(t,\varepsilon)$, $t \geq t_j$, is defined, see (10.101b), by

$$q_I(t,\varepsilon) := \int_{t_j}^{t} q(s,\varepsilon)ds, \tag{10.110}$$

problem (10.105a)–(10.105d) can be represented by the equivalent second-order initial value problem:

$$\begin{aligned} &F\big(p_D(\varepsilon), q(t,\varepsilon), \dot{q}(t,\varepsilon), \ddot{q}(t,\varepsilon)\big) \\ &\quad = u^{(j)}(t) + \varphi^{(j)}\big(t, q(t,\varepsilon) - q^{(j)}(t), q_I(t,\varepsilon) - q_I^{(j)}(t), \\ &\qquad\qquad \dot{q}(t,\varepsilon) - \dot{q}^{(j)}(t)\big) \qquad (10.111a) \\ &\dot{q}_I(t,\varepsilon) := q(t,\varepsilon) \qquad (10.111b) \end{aligned}$$

with

$$\begin{aligned} q(t_j,\varepsilon) &= q_j(\varepsilon) \qquad (10.111c) \\ \dot{q}(t_j,\varepsilon) &= \dot{q}_j(\varepsilon) \qquad (10.111d) \\ q_I(t_j,\varepsilon) &= 0. \qquad (10.111e) \end{aligned}$$

and

$$q_I^{(j)}(t) := \int_{t_j}^{t} q^{(j)}(s)ds. \tag{10.112}$$

10.7.2 Taylor Expansion with Respect to ε

Based on representation (10.111a)–(10.111e) of problem (10.105a)–(10.105d), we may again assume, cf. Assumption 10.1, that the solution $q = q(t,\varepsilon)$, $t \geq t_j$, $0 \leq \varepsilon \leq 1$, has continuous derivatives with respect to ε up to a certain order $\nu \geq 1$ for all $t \in [t_j, t_j + \Delta t_j]$, $0 \leq \varepsilon \leq 1$, with a certain $\Delta t_j > 0$.

Corresponding to (10.76a)–(10.76c), the actual trajectory of the robot can be represented then, see (10.109), (10.107), (10.104), by

$$q(t) = q\big(t, p_D, q_j, \dot{q}_j, u(\cdot)\big) = q(t, 1) = q(t, \varepsilon_0 + \Delta\varepsilon)$$
$$= q(t, \varepsilon_0) + \Delta q(t) = q^{(j)}(t) + \Delta q(t), \tag{10.113a}$$

with $\varepsilon_0 = 0$, $\Delta\varepsilon = 1$. Moreover, the expansion on the tracking error $\Delta q = \Delta q(t)$, $t \geq t_j$, is given by

$$\Delta q(t) = \sum_{l=1}^{\nu-1} \frac{1}{l!} d^l q(t)(\Delta\varepsilon)^l + \frac{1}{\nu!}\frac{\partial^\nu}{\partial\varepsilon^\nu} q(t, \vartheta)(\Delta\varepsilon)^\nu$$
$$= \sum_{l=1}^{\nu-1} \frac{1}{l!} d^l q(t) + \frac{1}{\nu!}\frac{\partial^\nu q}{\partial\varepsilon^\nu}(t, \vartheta), \quad t \geq t_j. \tag{10.113b}$$

Here, $\vartheta = \vartheta(t, \nu)$, $0 < \vartheta < 1$, and

$$d^l q(t) := \frac{\partial^l q}{\partial\varepsilon^l}(t, 0), \quad t \geq t_j, l = 1, 2, \ldots \tag{10.113c}$$

denote the l-th order differentials of $q = q(t, \varepsilon)$ with respect to ε at $\varepsilon = \varepsilon_0 = 0$. Differential equations for the differentials $d^l q(t)$, $l = 1, 2, \ldots$, may be obtained by successive differentiation of the initial value problem (10.105a)–(10.105d) with respect to ε at $\varepsilon = 0$.

10.7.3 The First-Order Differential dq

Corresponding to Sect. 10.6.2, we consider now the partial derivative in the Eqs. (10.111a)–(10.111e) with respect to ε. Let

$$K(p_D, q, \dot{q}, \ddot{q}) := F_q(p_D, q, \dot{q}, \ddot{q}) \tag{10.114a}$$
$$D(p_D, q, \dot{q}) := F_{\dot{q}}(p_D, q, \dot{q}, \ddot{q}) = h_{\dot{q}}(p_D, q, \dot{q}) \tag{10.114b}$$
$$Y(q, \dot{q}, \ddot{q}) := F_{p_D}(p_D, q, \dot{q}, \ddot{q}) \tag{10.114c}$$
$$M(p_D, q) := F_{\ddot{q}}(p_D, q, \dot{q}, \ddot{q}) \tag{10.114d}$$

denote again the Jacobians of the vector function $F = F(p_D, q, \dot{q}, \ddot{q})$ with respect to $q, \dot{q}, \ddot{q}$ and p_D. According to the linear parametrization property of robots we have, see (10.77e)

$$F(p_D, q, \dot{q}, \ddot{q}) = Y(q, \dot{q}, \ddot{q})p_D. \tag{10.114e}$$

Taking the partial derivative with respect to ε, from (10.111a)–(10.111e) we obtain the following equations, cf. (10.78a), (10.78b),

$$
\begin{aligned}
&Y\big(q(t,\varepsilon),\dot q(t,\varepsilon),\ddot q(t,\varepsilon)\big)\Delta p_D + K\big(p_D(\varepsilon),\dot q(t,\varepsilon),\ddot q(t,\varepsilon)\big)\frac{\partial q}{\partial \varepsilon}(t,\varepsilon)\\
&+ D\big(p_D(\varepsilon),q(t,\varepsilon),\dot q(t,\varepsilon),\ddot q(t,\varepsilon)\big)\frac{d}{dt}\frac{\partial q}{\partial \varepsilon}(t,\varepsilon) + M\big(p_D(\varepsilon),q(t,\varepsilon)\big)\frac{d^2}{dt^2}\frac{\partial q}{\partial \varepsilon}(t,\varepsilon)\\
&= \varphi_q^{(j)}\big(t,q(t,\varepsilon)-q^{(j)}(t),q_I(t,\varepsilon)-q_I^{(j)}(t),\dot q(t,\varepsilon)-\dot q^{(j)}(t)\big)\frac{\partial q}{\partial \varepsilon}(t,\varepsilon)\\
&+ \varphi_{q_I}^{(j)}\big(t,q(t,\varepsilon)-q^{(j)}(t),q_I(t,\varepsilon)-q_I^{(j)}(t),\dot q(t,\varepsilon)-\dot q^{(j)}(t)\big)\frac{\partial q_I}{\partial \varepsilon}(t,\varepsilon)\\
&+ \varphi_{\dot q}^{(j)}\big(t,q(t,\varepsilon)-q^{(j)}(t),q_I(t,\varepsilon)-q_I^{(j)}(t),\dot q(t,\varepsilon)-\dot q^{(j)}(t)\big)\frac{d}{dt}\frac{\partial q}{\partial \varepsilon}(t,\varepsilon).
\end{aligned}
\tag{10.115a}
$$

$$
\frac{d}{dt}\frac{\partial q_I}{\partial \varepsilon}(t,\varepsilon) = \frac{\partial q}{\partial \varepsilon}(t,\varepsilon) \tag{10.115b}
$$

$$
\frac{\partial q}{\partial \varepsilon}(t_j,\varepsilon) = \Delta q_j \tag{10.115c}
$$

$$
\frac{d}{dt}\frac{\partial q}{\partial \varepsilon}(t_j,\varepsilon) = \dot\Delta q_j \tag{10.115d}
$$

$$
\frac{\partial q_I}{\partial \varepsilon}(t,\varepsilon) = 0. \tag{10.115e}
$$

Putting now $\varepsilon = \varepsilon_0 = 0$, due to (10.107), (10.110), 10.113c) we obtain

$$
\begin{aligned}
&Y^{(j)}(t)\Delta p_D + K^{(j)}(t)dq(t) + D^{(j)}(t)\dot dq(t) + M^{(j)}(t)\ddot dq(t)\\
&= \varphi_q^{(j)}(t,0,0,0)dq(t) + \varphi_{q_I}^{(j)}(t,0,0,0)\int_{t_j}^{t} dq(s)ds + \varphi_{\dot q}^{(j)}(t,0,0,0)\dot dq(t), t \ge t_j
\end{aligned}
\tag{10.116a}
$$

$$
dq(t_j) = \Delta q_j \tag{10.116b}
$$

$$
\dot dq(t_j) = \dot\Delta q_j, \tag{10.116c}
$$

where

$$
\dot dq := \frac{d}{dt}dq, \quad \ddot dq := \frac{d^2}{dt^2}dq. \tag{10.117a}
$$

From (10.115b), (10.115e) we obtain

$$
\frac{\partial q_I}{\partial \varepsilon}(t,0) = \int_{t_j}^{t} dq(s)ds, \tag{10.117b}
$$

which is already used in (10.116a). Moreover, the matrices $Y^{(j)}$, $K^{(j)}$, $D^{(j)}$ and $M^{(j)}$ are defined as in Eqs. (10.79e)–(10.79g) of Sect. 10.6.2, hence,

$$Y^{(j)}(t) := Y\big(q^{(j)}(t), \dot{q}^{(j)}(t)\ddot{q}^{(j)}(t)\big) \tag{10.118a}$$
$$K^{(j)}(t) := K\big(\overline{p}_D^{(j)}, q^{(j)}(t), \dot{q}^{(j)}(t), \ddot{q}^{(j)}(t)\big) \tag{10.118b}$$
$$D^{(j)}(t) := D\big(\overline{p}_D^{(j)}, q^{(j)}(t), \dot{q}^{(j)}(t)\big) \tag{10.118c}$$
$$M^{(j)}(t) := M\big(\overline{p}_D^{(j)}, q^{(j)}(t)\big), t \geq t_j, \tag{10.118d}$$

see also (10.114a)–(10.114d).

Multiplying now (10.116a) with the inverse $M^{(j)}(t)^{-1}$ of $M^{(j)}(t)$ and rearranging terms, we get

$$\begin{aligned}\ddot{dq}(t) &+ M^{(j)}(t)^{-1}\big(D^{(j)}(t) - \varphi_{\dot{q}}^{(j)}(t,0,0,0)\big)\dot{dq}(t)\\ &+ M^{(j)}(t)^{-1}\big(K^{(j)}(t) - \varphi_{q}^{(j)}(t,0,0,0)\big)dq(t)\\ &- M^{(j)}(t)^{-1}\varphi_{q_I}^{(j)}(t,0,0,0)\int_{t_j}^{t} dq(s)ds = -M^{(j)}(t)^{-1}Y^{(j)}\Delta p_D, t \geq t_j\end{aligned} \tag{10.119}$$

with the initial conditions (10.116b), (10.116b).

For given matrices K_d, K_p, K_i to be selected later on, the unknown Jacobians $\varphi_{\dot{q}}^{(j)}$, $\varphi_q^{(j)}$ and $\varphi_{q_I}^{(j)}$ are defined now by the equations

$$M^{(j)}(t)^{-1}\big(D^{(j)}(t) - \varphi_{\dot{q}}^{(j)}(t,0,0,0)\big) = K_d \tag{10.120a}$$
$$M^{(j)}(t)^{-1}\big(K^{(j)}(t) - \varphi_{q}^{(j)}(t,0,0,0)\big) = K_p \tag{10.120b}$$
$$M^{(j)}(t)^{-1}\big(-\varphi_{q_I}^{(j)}(t,0,0,0)\big) = K_i. \tag{10.120c}$$

Thus, we have

$$\varphi_q^{(j)}(t,0,0,0) = K^{(j)}(t) - M^{(j)}(t)K_p \tag{10.121a}$$
$$\varphi_{q_I}^{(j)}(t,0,0,0) = -M^{(j)}(t)K_i \tag{10.121b}$$
$$\varphi_{\dot{q}}^{(j)}(t,0,0,0) = D^{(j)}(t) - M^{(j)}(t)K_d. \tag{10.121c}$$

Putting (10.120a)–(10.120c) into system (10.119), we find

$$\ddot{dq}(t) + K_d\dot{dq}(t) + K_p dq(t) + K_i\int_{t_j}^{t} dq(s)ds = \psi^{(j,1)}(t), \quad t \geq t_j, \tag{10.122a}$$

with

$$dq(t_j) = \Delta q_j \tag{10.122b}$$
$$\dot{dq}(t_j) = \dot{\Delta} q_j, \tag{10.122c}$$

where $\psi^{(j,i1}(t)$ is given, cf. (10.84d), by

$$\psi^{(j,1)}(t) := -M^{(j)}(t)^{-1} Y^{(j)}(t) \Delta p_D. \tag{10.122d}$$

If K_p, K_d and K_i are fixed matrices, then by differentiation of the integro-differential equation (10.122a) with respect to time t we obtain the following third-order system of linear differential equations

$$\dddot{dq}(t) + K_d \ddot{dq}(t) + K_p \dot{dq}(t) + K_i dq(t) = \dot{\psi}^{(j,1)}(t), \quad t \geq t_j, \tag{10.123a}$$

for the first-order tracking error term $dq = dq(t), t = t_j$. Moreover, the initial conditions read, see (10.122b), (10.122c),

$$dq(t_j) = \Delta q_j \tag{10.123b}$$

$$\dot{dq}(t_j) = \dot{\Delta} q_j \tag{10.123c}$$

$$\ddot{dq}(t_j) = \ddot{\Delta} q_j := \ddot{q}_j - \ddot{q}^{(j)}, \tag{10.123d}$$

where $\ddot{q}_j := \ddot{q}(t_j)$, see (10.103a)–(10.103c).

Corresponding to (10.84a)–(10.84c), the system (10.122a) of third-order linear differential equations can be converted easily into the following system of first-order differential equations

$$\dot{dz}(t) = A dz(t) + \begin{pmatrix} 0 \\ 0 \\ \dot{\psi}^{(j,1)}(t) \end{pmatrix}, \quad t \geq t_j, \tag{10.124a}$$

where

$$A := \begin{pmatrix} 0 & I & 0 \\ 0 & 0 & I \\ -K_i & -K_p & -K_d \end{pmatrix}, \tag{10.124b}$$

$$dz(t_j) := \begin{pmatrix} \Delta q_j \\ \dot{\Delta} q_j \\ \ddot{\Delta} q_j \end{pmatrix} = \Delta z_j \tag{10.124c}$$

and

$$dz(t) := \begin{pmatrix} dq(t) \\ \dot{dq}(t) \\ \ddot{dq}(t) \end{pmatrix}. \tag{10.124d}$$

With the fundamental matrix $\Phi(t, \tau) := e^{A(t-\tau)}$, $t \geq \tau$, the solution of (10.124a)–(10.124d) reads

$$dz(t) = e^{A(t-t_j)}\Delta z_j + \int_{t_j}^{t} e^{A(t-\tau)} \begin{pmatrix} 0 \\ 0 \\ \dot{\psi}^{(j,1)}(\tau) \end{pmatrix} d\tau, t \geq t_j, \tag{10.125a}$$

where, see (10.122d),

$$\dot{\psi}^{(j,1)}(t) = -\frac{d}{dt}\Big(M^{(j)}(t)^{-1}Y^{(j)}(t)\Big)\Delta p_D. \tag{10.125b}$$

Because of

$$\Delta p_D = p_D(\omega) - \overline{p}_D^{(j)} = p_D(\omega) - E\Big(p_D(\omega)|\mathfrak{A}_{t_j}\Big),$$

see (10.103c), for the conditional mean first-order error term $E\Big(dz(t)|\mathfrak{A}_{t_j}\Big)$ from (10.125a), (10.125b) we get

$$E\Big(dz(t)|\mathfrak{A}_{t_j}\Big) = e^{A(t-t_j)}E(\Delta z_j|\mathcal{A}_{t_j}), \quad t \geq t_j. \tag{10.125c}$$

Obviously, the properties of the first-order error terms $dz(t)$, $E(dz(t)|\mathfrak{A}_{t_j})$, resp., $t \geq t_j$, or the stability properties of the 1st order system (10.124a)–(10.124d) depend on the eigenvalues of the matrix A.

10.7.3.1 Diagonalmatrices K_p, K_d, K_i

Supposing here, cf. Sect. 10.5, that K_p, K_d, K_i are diagonal matrices

$$K_p = (\gamma_{pk}\delta_{k\kappa}),\ K_d = (\gamma_{dk}\delta_{k\kappa}),\ K_i = (\gamma_{ik}\delta_{k\kappa}) \tag{10.126}$$

with diagonal elements γ_{pk}, γ_{dk}, γ_{ik}, resp., $k = 1, \ldots, n$, system (10.123a) is divided into the separated ordinary differential equations

$$\dddot{d}q_k(t) + \gamma_{dk}\ddot{d}q_k(t) + \gamma_{pk}\dot{d}q_k(t) + \gamma_{ik}dq_k(t) = \dot{\psi}_k^{(j,1)}(t), \quad t \geq t_j, \tag{10.127a}$$

$k = 1, 2, \ldots, n$. The related homogeneous differential equations read

$$\dddot{d}q_k + \gamma_{dk}\ddot{d}q_k + \gamma_{pk}\dot{d}q_k + \gamma_{ik}dq_k = 0, \tag{10.127b}$$

$k = 1, \ldots, n$, which have the characteristic equations

$$\lambda^3 + \gamma_{dk}\lambda^2 + \gamma_{pk}\lambda + \gamma_{ik} =: p_k(\lambda) = 0, \quad k = 1, \ldots, n, \tag{10.127c}$$

with the polynomials $p_k = p_k(\lambda)$, $k = 1, \ldots, n$, of degree $= 3$.

A system described by the homogeneous differential equation (10.127b) is called **uniformly (asymptotic) stable** if

$$\lim_{t\to\infty} dq_k(t) = 0 \tag{10.128a}$$

for arbitrary initial values (10.123b)–(10.123d). It is well known that property (10.128a) holds if

$$Re(\lambda_{kl}) < 0, \quad l = 1, 2, 3, \tag{10.128b}$$

where $Re(\lambda_{kl})$ denotes the real part of the zeros $\lambda_{k1}, \lambda_{k2}, \lambda_{k3}$ of the characteristic equation (10.127c). According to the **Hurwitz criterion**, a necessary and sufficient condition for (10.128b) is the set of inequalities

$$\det(\gamma_{dk}) > 0 \tag{10.129a}$$

$$\det\begin{pmatrix} \gamma_{dk} & 1 \\ \gamma_{ik} & \gamma_{pk} \end{pmatrix} = \gamma_{dk}\gamma_{pk} - \gamma_{ik} > 0 \tag{10.129b}$$

$$\det\begin{pmatrix} \gamma_{dk} & 1 & 0 \\ \gamma_{ik} & \gamma_{pk} & \gamma_{dk} \\ 0 & 0 & \gamma_{ik} \end{pmatrix} = \gamma_{ik}(\gamma_{dk}\gamma_{pk} - \gamma_{ik}) > 0. \tag{10.129c}$$

Note that

$$H_{3k} := \begin{pmatrix} \gamma_{dk} & 1 & 0 \\ \gamma_{ik} & \gamma_{pk} & \gamma_{dk} \\ 0 & 0 & \gamma_{ik} \end{pmatrix} \tag{10.129d}$$

is the so-called **Hurwitz matrix** of (10.127b).

Obviously, from (10.129a)–(10.129c) we now obtain this result:

Theorem 10.8 *The system represented by the homogeneous third-order linear differential equation (10.127b) is uniformly (asymptotic) stable if the (feedback) coefficients $\gamma_{pk}, \gamma_{dk}, \gamma_{ik}$ are selected such that*

$$\gamma_{pk} > 0, \gamma_{dk} > 0, \gamma_{ik} > 0 \tag{10.130a}$$

$$\gamma_{dk}\gamma_{pk} > \gamma_{ik}. \tag{10.130b}$$

10.7.3.2 Mean Absolute First-Order Tracking Error

Because of $E\left(\Delta p_D(\omega)|\mathfrak{A}_{t_j}\right) = 0$ and the representation (10.125b) of $\dot{\psi}^{(j,1)} = \dot{\psi}^{(j,1)}(t)$, corresponding to (10.85b) we have

$$E\left(\dot{\psi}^{(j,1)}(t)|\mathfrak{A}_{t_j}\right) = 0, \quad t \geq t_j. \tag{10.131a}$$

Hence, (10.125a) yields

$$E\Big(dz(t)|\mathfrak{A}_{t_j}\Big) = e^{A(t-t_j)}E\Big(\Delta z_j|\mathfrak{A}_{t_j}\Big), \tag{10.131b}$$

where

$$E\Big(\Delta z_j|\mathfrak{A}_{t_j}\Big) = E\Big(z(t_j)|\mathcal{A}_{t_j}\Big) - \bar{z}_j \tag{10.131c}$$

with, cf. (10.106a), (10.123d),

$$\bar{z}_j = \begin{pmatrix} \overline{q}_j \\ \overline{\dot{q}}_j \\ \overline{\ddot{q}}^{(j)} \end{pmatrix}, z(t_j) = \begin{pmatrix} q_j \\ \dot{q}_j \\ \ddot{q}_j \end{pmatrix}. \tag{10.131d}$$

The matrices K_p, K_i, K_d in the definition (10.120a)–(10.120c) or (10.121a)–(10.121c) of the Jacobians $\varphi_q^{(j)}(t,0,0,0)$, $\varphi_{q_I}^{(j)}(t,0,0,0)$, $\varphi_{\dot{q}}^{(j)}(t,0,0,0)$ of the feedback control law $\Delta u^{(j)}(t) = \varphi^{(j)}(t, \Delta q, \Delta q_I, \dot{\Delta} q)$ can be chosen now, see Theorem 10.8, such that the fundamental matrix $\Phi(t,\tau) = e^{A(t-\tau)}$, $t \geq \tau$, is exponentially stable, hence,

$$\|\Phi(t,\tau)\| \leq a_0 e^{-\lambda_0(t-\tau)}, \quad t \geq \tau, \tag{10.132}$$

with constants $a_0 > 0$, $\lambda_0 > 0$, see also (10.86a).

Considering the Euclidean norm $\|dz(t)\|$ of the first-order error term $dz(t)$, from (10.125a), (10.125b) and with (10.132) we obtain

$$\begin{aligned} \|dz(t)\| &\leq \|e^{A(t-t_j)}\Delta z_j\| + \left\| \int_{t_j}^{t} e^{A(t-\tau)} \begin{pmatrix} 0 \\ 0 \\ \dot{\psi}^{(j,1)}(\tau) \end{pmatrix} d\tau \right\| \\ &\leq a_0 e^{-\lambda_0(t-t_j)}\|\Delta z_j\| + \int_{t_j}^{t} \left\| e^{A(t-\tau)} \begin{pmatrix} 0 \\ 0 \\ \dot{\psi}^{(j,1)}(\tau) \end{pmatrix} \right\| d\tau \\ &\leq a_0 e^{-\lambda_0(t-t_j)}\|\Delta z_j\| + \int_{t_j}^{t} a_0 e^{-\lambda_0(t-\tau)}\|\dot{\psi}^{(j,1)}(\tau)\| d\tau. \end{aligned} \tag{10.133}$$

Taking the conditional expectation in (10.133), we find

$$\begin{aligned} E\Big(\|dz(t)\|\Big|\mathfrak{A}_{t_j}\Big) \leq & a_0 e^{-\lambda_0(t-t_j)} E\Big(\|\Delta z_j\|\Big|\mathfrak{A}_{t_j}\Big) \\ & + a_0 \int_{t_j}^{t} e^{-\lambda_0(t-\tau)} E\Big(\|\dot{\psi}^{(j,1)}(\tau)\|\Big|\mathfrak{A}_{t_j}\Big) d\tau. \end{aligned} \tag{10.134}$$

Applying Jensen's inequality

$$\sqrt{EX(\omega)} \geq E\sqrt{X(\omega)},$$

where $X = X(\omega)$ is a nonnegative random variable, with (10.125b) we get, cf. (10.87d), (10.87e),

$$\begin{aligned} E\Big(\|\dot{\psi}^{(j,1)}(\tau)\|\Big|\mathfrak{A}_{t_j}\Big) &= E\Big(\sqrt{\|\dot{\psi}^{(j,1)}(\tau)\|^2}\Big|\mathfrak{A}_{t_j}\Big) \\ &\leq \sqrt{E\Big(\|\dot{\psi}^{(j,1)}(\tau)\|^2\Big|\mathcal{A}_{t_j}\Big)} \\ &\leq \left\|\frac{d\Big(M^{(j)}(t)^{-1}Y^{(j)}(t)\Big)}{dt}(\tau)\right\| \sqrt{\text{var}\,\Big(p_D(\cdot)\Big|\mathcal{A}_{t_j}\Big)}, \end{aligned} \tag{10.135a}$$

where

$$\text{var}\,\Big(p_D(\cdot)\Big|\mathfrak{A}_{t_j}\Big) := E\Big(\|\Delta p_D(\omega)\|^2\Big|\mathfrak{A}_{t_j}\Big). \tag{10.135b}$$

In the following we study the inhomogeneous term $\dot{\psi}^{(j,1)}(t)$ of the third-order linear differential equation (10.123a) in more detail. According to (10.122d) we have

$$\begin{aligned} \dot{\psi}^{(j,1)}(t) &= -\frac{d}{dt}\Big(M^{(j)}(t)^{-1}Y^{(j)}(t)\Big)\Delta p_D \\ &= -\left(\frac{d}{dt}M^{(j)}(t)^{-1}\right)Y^{(j)}(t)\Delta p_D \\ &\quad - M^{(j)}(t)^{-1}\left(\frac{d}{dt}Y^{(j)}(t)\right)\Delta p_D. \end{aligned} \tag{10.136a}$$

and therefore

$$\begin{aligned} \|\dot{\psi}^{(j,1)}(t)\| \leq &\left\|\frac{d}{dt}M^{(j)}(t)^{-1}\right\| \cdot \|Y^{(j)}(t)\Delta p_D\| \\ &+ \|M^{(j)}(t)^{-1}\| \cdot \left\|\Big(\dot{Y}^{(j)}(t)\Big)\Delta p_D\right\|. \end{aligned} \tag{10.136b}$$

Now, according to (10.20a), (10.20b), (10.83a)–(10.83c) it holds

$$u^{(j)}(t;\Delta p_D) := Y^{(j)}(t)\Delta p_D = u_e\Big(\Delta p_D, s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\Big), t \geq t_j. \tag{10.137a}$$

Thus, we find

$$\begin{aligned}\dot{u}^{(j)}(t;\Delta p_D) &= \left(\dot{Y}^{(j)}(t)\right)\Delta p_D \\ &= \frac{d}{dt}u_e\left(\Delta p_D, s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\right) \\ &= \frac{\partial u_e}{\partial s}\left(\Delta p_D, s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\right)\cdot\frac{d}{dt}s^{(j)}(t) \\ &= \frac{\partial u_e}{\partial s}\left(\Delta p_D, s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\right)\cdot\sqrt{\beta^{(j)}\left(s^{(j)}(t)\right)}, \qquad (10.137b)\end{aligned}$$

see (10.18c), (10.50b). Note that

$$u^{(j)}(t;\Delta p_D), \dot{u}^{(j)}(t;\Delta p_D)$$

are linear with respect to $\Delta p_D = p_D - E\left(p_D(\omega)|\mathfrak{A}_{t_j}\right)$.

From (10.136a), (10.136b) and (10.137a), (10.137b) we get

$$\begin{aligned}E\left(\|\dot{\psi}^{(j,1)}(t)\|\Big|\mathfrak{A}_{t_j}\right) &\le \left\|\frac{d}{dt}M^{(j)}(t)^{-1}\right\| E\left(\|Y^{(j)}(t)\Delta p_D\|\Big|\mathfrak{A}_{t_j}\right) \\ &\quad + \|M^{(j)}(t)^{-1}\| E\left(\left\|\left(\dot{Y}^{(j)}(t)\right)\Delta p_D\right\|\Big|\mathfrak{A}_{t_j}\right) \\ &= \left\|\frac{d\left(M^{(j)}(t)^{-1}\right)}{dt}\right\|\cdot E\left(\|u(j)(t;\Delta p_D)\|\Big|\mathfrak{A}_{t_j}\right) \\ &\quad + \|M^{(j)}(t)^{-1}\|\cdot E\left(\|\dot{u}^{(j)}(t;\Delta p_D\|\Big|\mathfrak{A}_{t_j}\right) \\ &= \left\|\frac{d\left(M^{(j)}(t)^{-1}\right)}{dt}\right\|\cdot E\left(\left\|u_e\left(\Delta p_D, s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\right)\right\|\Big|\mathfrak{A}_{t_j}\right) \\ &+ \|M^{(j)}(t)^{-1}\|\cdot E\left(\left\|\frac{\partial u_e}{\partial s}\left(\Delta p_D, s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\right)\sqrt{\beta^{(j)}\left(s^{(j)}(t)\right)}\right\|\Big|\mathfrak{A}_{t_j}\right). \qquad (10.138a)\end{aligned}$$

Using again Jensen's inequality, from (10.138a) we obtain

$$\begin{aligned}&E\left(\|\dot{\psi}^{(j,1)}(t)\|\Big|\mathfrak{A}_{t_j}\right) \\ &\le \left\|\frac{d\left(M^{(j)}(t)^{-1}\right)}{dt}\right\|\sqrt{\mathrm{var}\left(u_e\left(p_D(\cdot), s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\right)\Big|\mathfrak{A}_{t_j}\right)} \\ &+ \|M^{(j)}(t)^{-1}\|\sqrt{\mathrm{var}\left(\frac{\partial u_e}{\partial s}\left(p_D(\cdot), s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\right)\Big|\mathfrak{A}_{t_j}\right)\beta^{(j)}\left(s^{(j)}(t)\right)}, \qquad (10.138b)\end{aligned}$$

where

$$\begin{aligned}&\operatorname{var}\Big(u_e\big(p_D(\cdot), s^{(j)}(t); q_0^{(j)}(\cdot), \beta^{(j)}(\cdot)\big)\Big|\mathfrak{A}_{t_j}\Big)\\ &:= E\Big(\Big\|u_e\big(\Delta p_D(\omega), s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\big)\Big\|^2\Big|\mathfrak{A}_{t_j}\Big),\end{aligned} \tag{10.138c}$$

$$\begin{aligned}&\operatorname{var}\left(\frac{\partial u_e}{\partial s}\Big(p_D(\cdot), s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\Big)\sqrt{\beta^{(j)}\Big(s^{(j)}(t)\Big)}\Big|\mathfrak{A}_{t_j}\right)\\ &= E\left(\Big\|\frac{\partial u_e}{\partial s}\Big(\Delta p_D(\omega), s^{(j)}(t); q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\Big)\Big\|^2\Big|\mathfrak{A}_{t_j}\right)\beta^{(j)}\Big(s^{(j)}(t)\Big).\end{aligned} \tag{10.138d}$$

According to the representation (10.125a), (10.125b) of the first-order tracking error form $dz = dz(t), t \geq t_j$, the behavior of $dz(t)$ is determined mainly by the system matrix A and the "inhomogeneous term" $\dot{\psi}^{(j,1)}(t), t \geq t_j$. Obviously, this term plays the same role as the expression $\psi^{(j,1)}(t), t \geq t_j$, in the representation (10.85a) for the first-order error term in case of PD-controllers.

However, in the present case of PID-control the error estimates (10.138a)–(10.138d) show that for a satisfactory behavior of the first-order error term $dz = dz(t), t \geq t_j$, besides the control constraints (10.9a)–(10.9c), (10.22a)–(10.22c), (10.31a)–(10.31f) for

$$u(t) = u_e\Big(p_D, s; q_e(.), \beta(.)\Big) \text{ with } s = s(t), \tag{10.139a}$$

here, also corresponding constraints for the input rate, i.e., the time derivative of the control

$$\dot{u}(t) = \frac{\partial u_e}{\partial s}\Big(p_D, s; q_e(.), \beta(.)\Big)\sqrt{\beta(s)} \text{ with } s = s(t) \tag{10.139b}$$

are needed!

The above results can be summarized by the following theorem:

Theorem 10.9 *Suppose that the matrices K_p, K_i, K_d are selected such that the fundamental matrix $\Phi(t, \tau) = e^{A(t-\tau)}, t \geq \tau$, is exponentially stable (cf. Theorem 10.8). Then, based on the definition (10.121a)–(10.121c) of the linear approximation of the PID-controller, the following properties hold:*

(a) ***Asymptotic local stability in the mean***

$$E\big(dz(t)\big|\mathfrak{A}_{t_j}\big) \to 0, \quad t \to \infty \tag{10.140a}$$

(b) ***Mean absolute first-order tracking error***

$$E\left(\|dz(t)\|\middle|\mathfrak{A}_{t_j}\right) \le a_0 e^{-\lambda_0(t-t_j)} E\left(\|\Delta z_j\|\middle|\mathfrak{A}_{t_j}\right) + a_0 \int_{t_j}^{t} e^{-\lambda_0(t-\tau)} \left\| \frac{d\left(M^{(j)}(t)^{-1}\right)}{dt}(\tau) \right\| \sigma_{u_e}^{(j)}\left(s^{(j)}(\tau)\right) d\tau + a_0 \int_{t_j}^{t} e^{-\lambda_0(t-\tau)} \left\| M^{(j)}(\tau)^{-1} \right\| \sigma_{\frac{\partial u_e}{\partial s}}^{(j)}\left(s^{(j)}(\tau)\right) \sqrt{\beta^{(j)}\left(s^{(j)}(\tau)\right)} d\tau, \; t \ge t_j \tag{10.140b}$$

with

$$\sigma_{u_e}^{(j)}(s) := \sqrt{\mathrm{var}\left(u_e\left(p_D(.), s; q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\right)\middle|\mathfrak{A}_{t_j}\right)}, \tag{10.140c}$$

$$\sigma_{\frac{\partial u_e}{\partial s}}^{(j)}(s) := \sqrt{\mathrm{var}\left(\frac{\partial u_e}{\partial s}\left(p_D(\cdot), s; q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\right)\middle|\mathfrak{A}_{t_j}\right)}, \tag{10.140d}$$

$s \ge s_j$. *Moreover,*

$$E\left(\|dz(t)\|\middle|\mathfrak{A}_{t_j}\right) \le a_0 e^{-\lambda_0(t-t_j)} E\left(\|\Delta z_j\|\middle|\mathfrak{A}_{t_j}\right) + a_0 \left(\int_{t_j}^{t} e^{-\lambda_0(t-\tau)} \left\| \frac{d\left(M^{(j)}(t)^{-1} Y^{(j)}(t)\right)}{dt}(\tau) \right\| d\tau \right) \sigma_{p_D}^{(j)}, \; t \ge t_j, \tag{10.140e}$$

where

$$\sigma_{p_D}^{(j)} := \sqrt{\mathrm{var}\left(p_D(.)\middle|\mathfrak{A}_{t_j}\right)}. \tag{10.140f}$$

Using the t-s-transformation $s = s^{(j)}(\tau)$, $\tau \ge t_j$, the time-integrals in (10.140b) can be represented also in the following form:

$$\int_{t_j}^{t} e^{-\lambda_0(t-\tau)} \left\| \frac{d\left(M^{(j)}(t)^{-1}\right)}{dt}(\tau) \right\| \sigma_{u_e}^{(j)}\left(s^{(j)}(\tau)\right) d\tau = \int_{s_j}^{s^{(j)}(t)} e^{-\lambda_0\left(t-t^{(j)}(s)\right)} \left\| \frac{d\left(M^{(j)}(t)^{-1}\right)}{dt}\left(t^{(j)}(s)\right) \right\| \frac{\sigma_{u_e}^{(j)}(s)}{\sqrt{\beta^{(j)}(s)}} ds, \tag{10.141a}$$

$$\int_{t_j}^{t} e^{-\lambda_0(t-\tau)} \left\| M^{(j)}(\tau)^{-1} \right\| \sigma^{(j)}_{\frac{\partial u_e}{\partial s}}\left(s^{(j)}(\tau)\right) \sqrt{\beta^{(j)}\left(s^{(j)}(\tau)\right)} d\tau$$

$$= \int_{s_j}^{s^{(j)}(t)} e^{-\lambda_0\left(t-t^{(j)}(s)\right)} \left\| M^{(j)}\left(t^{(j)}(s)\right)^{-1} \right\| \sigma^{(j)}_{\frac{\partial u_e}{\partial s}}(s) ds, \tag{10.141b}$$

where $\tau = t^{(j)}(s)$, $s \geq s_j$, denotes the inverse of $s = s^{(j)}(\tau)$, $\tau \geq t_j$.

10.7.3.3 Minimality or Boundedness Properties

Several terms in the above estimates of the first-order tracking error $dz = dz(t)$, $t \geq t_j$, are influenced by means of (OSTP) as shown in the following:

(i) **Optimal velocity profile** $\beta^{(j)}$
For minimum-time and related substitute problems the total runtime

$$\int_{s_j}^{s^{(j)}(t)} \frac{1}{\sqrt{\beta^{(j)}(s)}} ds \leq \int_{s_j}^{s_f} \frac{1}{\sqrt{\beta^{(j)}(s)}} ds = t_f^{(j)} - t_j \tag{10.142}$$

is minimized by (OSTP).

(ii) **Properties of the coefficient** λ_0
According to Theorems 10.8 and 10.9 the matrices K_p, K_i, K_d can be selected such that real parts $\mathrm{Re}\left(\lambda_{kl}\right)$, $k = 1, \ldots, n$, $l = 1, 2, 3$, of the eigenvalues λ_{kl}, $k = 1, \ldots, n$, $l = 1, 2, 3$, of the matrix A, cf. (10.124b), are negative, see (10.128b), (10.129a)–(10.129d). Then, the decisive coefficient $\lambda_0 > 0$ in the norm estimate (10.132) of the fundamental matrix $\Phi(t, \tau)$, $t \geq \tau$, can be selected such that

$$0 < \lambda_0 < -\max_{\substack{1 \leq k \leq n \\ l=1,2,3}} \mathrm{Re}(\lambda_{kl}). \tag{10.143}$$

(iii) **Chance constraints for the input**
With certain lower and upper vector bounds $u^{\min} \leq u^{\max}$ one usually has the input or control constraints, cf. (10.9a),

$$u^{\min} \leq u(t) \leq u^{\max}, \quad t \geq t_j.$$

In the following we suppose that the bounds $u^{\min}$, $u^{\max}$ are given deterministic vectors. After the transformation

$$s = s^{(j)}(t), \quad t \geq t_j,$$

from the time domain $[t_j, t_f^{(j)}]$ to the s-domain $[s_j, s_f]$, see (10.50b), due to (10.22a) we have the stochastic constraint for $\big(q_e(\cdot), \beta(\cdot)\big)$:

$$u^{\min} \le u_e\big(p_D(\omega), s; q_e(\cdot), \beta(\cdot)\big) \le u^{\max}, \quad s_j \le s \le s_f.$$

Demanding that the above constraint holds at least with the probability α_u, we get, cf. (10.31a), the probabilistic constraint

$$P\Big(u^{\min} \le u_e\big(p_D(\omega), s; q_e(\cdot), \beta(\cdot)\big) \le u^{\max}\Big|\mathfrak{A}_{t_j}\Big) \ge \alpha_u, s_j \le s \le s_f. \tag{10.144a}$$

Defining again, see (10.31d),

$$u^c := \frac{u^{\min} + u^{\max}}{2}, \quad \rho_u := \frac{u^{\max} - u^{\min}}{2},$$

by means of Tschebyscheff-type inequalities, the chance constraint (10.144a) can be guaranteed, see (10.31e), (10.31f), (10.88d), by the condition

$$E\Big(\big\|u_e\big(p_D(\omega), s; q_e(.), \beta(.)\big) - u^c\big\|^2\Big|\mathfrak{A}_{t_j}\Big) \le (1-\alpha_u)\min_{1\le k\le n}\rho_{u_k}^2, \ s_j \le s \le s_f. \tag{10.144b}$$

According to the definition (10.140c) of $\sigma_{u_e}^{(j)}(s)$, inequality (10.144b) is equivalent to

$$\sigma_{u_e}^{(j)}(s)^2 + \big\|\overline{u}_e^{(j)}(s) - u^c\big\|^2 \le (1-\alpha_u)\min_{1\le k\le n}\rho_{u_k}^2, \ s_j \le s \le s_f, \tag{10.145a}$$

where

$$\begin{aligned}\overline{u}_e^{(j)} &:= E\Big(u_e\big(p_D(\omega), s; q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\big)\Big|\mathfrak{A}_{t_j}\Big)\\ &= u_e\Big(\overline{p}_D^{(j)}, s; q_e^{(j)}(\cdot), \beta^{(j)}(\cdot)\Big).\end{aligned} \tag{10.145b}$$

Hence, the sufficient condition (10.145a) for the reliability constraint (10.144a) yields the variance constraint

$$\sigma_{u_e}^{(j)}(s)^2 \le (1-\alpha_u)\min_{1\le k\le n}\rho_{u_k}^2, \quad s_j \le s \le s_f. \tag{10.145c}$$

(iv) **Minimum force and moment**

According to the different performance functions $J\big(u(.)\big)$ mentioned after the definition (10.6), in case of minimum expected force and moment we have, using transformation formula (10.20a),

$$
\begin{aligned}
E\Big(J\big(u(\cdot)\big)\Big|\mathfrak{A}_{t_j}\Big) &= E\Big(\int_{t_j}^{t_f^{(j)}} \|u(t)\|^2 dt\Big|\mathfrak{A}_{t_j}\Big) \\
&= \int_{s_j}^{s_f} E\Big(\big\|u_e\big(p_D(\omega), s; q_e(\cdot), \beta(\cdot)\big\|^2\Big|\mathfrak{A}_{t_j}\Big)\frac{ds}{\sqrt{\beta(s)}} \\
&= \int_{s_j}^{s_f} \Big(E\Big(\big\|u_e\big(p_D(\omega), s; q_e(\cdot), \beta(\cdot) - u_e\big(\overline{p}_D^{(j)}, s; q_e(\cdot), \beta(\cdot)\big\|^2\Big|\mathfrak{A}_{t_j}\Big) \\
&\quad + \big\|u_e\big(\overline{p}_D^{(j)}, s; q_e(\cdot), \beta(\cdot)\big)\big\|^2\Big)\frac{ds}{\sqrt{\beta(s)}} \\
&= \int_{s_j}^{s_f} \Big(\sigma_{u_e}^2(s) + \big\|u_e\big(\overline{p}_D^{(j)}, s; q_e(\cdot), \beta(\cdot)\big\|^2\Big)\frac{ds}{\sqrt{\beta(s)}}, \qquad (10.146a)
\end{aligned}
$$

where

$$
\begin{aligned}
\sigma_{u_e}^2(s) :&= E\Big(\Big\|u_e\big(p_D(\omega), s; q_e(\cdot), \beta(\cdot) - u_e\big(\overline{p}_D^{(j)}, s; q_e(\cdot), \beta(\cdot)\big)\Big\|^2\Big|\mathfrak{A}_{t_j}\Big) \\
&= E\Big(\Big\|u_e\big(\Delta p_D(\omega), s; q_e(\cdot), \beta(\cdot)\Big\|^2\Big|\mathfrak{A}_{t_j}\Big). \qquad (10.146b)
\end{aligned}
$$

Hence, (OSTP) yields the following **minimum property**:

$$
\int_{s_j}^{s_f} \Big(\sigma_{u_e}^{(j)}(s)^2 + \big\|\overline{u}_e^{(j)}(s)\big\|^2\Big)\frac{ds}{\sqrt{\beta^{(j)}(s)}} = \min_{\substack{q_e(\cdot),\beta(\cdot)\\ s.t.(10a-c),\\ (10a-d)}} E\Big(\int_{t_j}^{t_f^{(j)}} \|u(t)\|^2 dt\Big|\mathfrak{A}_{t_j}\Big), \qquad (10.147)
$$

where $\overline{u}_e^{(j)}(s)$ is defined by (10.145b).

(v) **Decreasing stochastic uncertainty**
According to (10.133), (10.134) and (10.135a), (10.135b) we have

$$
\begin{aligned}
E\Big(\|dz(t)\|\mathfrak{A}_{t_j}\Big) &\le \alpha_0 e^{-\lambda_0(t-t_j)} E\Big(\|\Delta z_j\|\Big|\mathfrak{A}_{t_j}\Big) \\
&\quad + \Big(a_o \int_{t_j}^{t} e^{-\lambda_0(t-\tau)} \left\| \frac{d\Big(M^{(j)}(t)^{-1} Y^{(j)}(t)\Big)}{dt}(\tau) \right\| d\tau\Big)\sqrt{\mathrm{var}\big(p_D(.)\big|\mathfrak{A}_{t_j}\big)}. \qquad (10.148)
\end{aligned}
$$

Thus, the mean absolute first-order tracking error can be decreased further by removing step by step the uncertainty about the vector $p_D = p_D(\omega)$ of dynamic

parameter. This is done in practice by a parameter identification procedure running parallel to control process of the robot.

(vi) **Chance constraints for the input rate**
According to the representation (10.139b) of the input or control rate we have

$$\dot{u}(t) = \frac{\partial u_e}{\partial s}\Big(p_D, s; q_e(\cdot), \beta(\cdot)\Big)\sqrt{\beta(s)}, \quad s = s(t).$$

From the input rate condition

$$\dot{u}^{\min} \le \dot{u}(t) \le \dot{u}^{\max}, t \ge t_j, \tag{10.149a}$$

with given, fixed vector bounds $\dot{u}^{\min} \le \dot{u}^{\max}$, for the input rate $\frac{\partial u_e}{\partial s}$ with respect to the path parameter $s \ge s_j$ we obtain the constraint

$$\frac{\dot{u}^{\min}}{\sqrt{\beta(s)}} \le \frac{\partial u_e}{\partial s}\big(p_D(\omega), s; q_e(\cdot), \beta(\cdot)\big) \le \frac{\dot{u}^{\max}}{\sqrt{\beta(s)}}, \quad s_j \le s \le s_f. \tag{10.149b}$$

If we require that the input rate condition (10.149a) holds at least with probability $\alpha_{\dot{u}}$, then corresponding to (10.144a) we get the chance constraint

$$P\left(\frac{\dot{u}^{\min}}{\sqrt{\beta(s)}} \le \frac{\partial u_e}{\partial s}\big(p_D(\omega), s; q_e(\cdot), \beta(\cdot)\big) \le \frac{\dot{u}^{\max}}{\sqrt{\beta(s)}}\middle|\mathfrak{A}_{t_j}\right) \ge \alpha_{\dot{u}}. \tag{10.150}$$

In the same way as in (iii), condition (10.150) can be guaranteed, cf. (10.140d), by

$$\sigma^{(j)}_{\frac{\partial u}{\partial s}}(s)^2 + \left\| \overline{\frac{\partial u_e}{\partial s}}^{(j)}(s) - \frac{\dot{u}^c}{\sqrt{\beta^{(j)}(s)}} \right\|^2 \le \Big(1 - \alpha_{\dot{u}}\Big)\frac{1}{\beta^{(j)}(s)} \min_{1\le k\le n} \rho^2_{\dot{u}_k}, \; s_j \le s \le s_f, \tag{10.151a}$$

where

$$\dot{u}^c := \frac{\dot{u}^{\min} + \dot{u}^{\max}}{2}, \quad \rho_{\dot{u}} := \frac{\dot{u}^{\max} - \dot{u}^{\min}}{2}, \tag{10.151b}$$

$$\overline{\frac{\partial u_e}{\partial s}}^{(j)} := \frac{\partial u_e}{\partial s}\Big(\overline{p}^{(j)}_D, s; q^{(j)}_e(\cdot), \beta^{(j)}(\cdot)\Big), \quad s \ge s_j. \tag{10.151c}$$

Hence, corresponding to (10.145c), here we get the following variance constraint:

$$\sigma^{(j)}_{\frac{\partial u}{\partial s}}(s)^2 \le \Big(1 - \alpha_{\dot{u}}\Big)\frac{1}{\beta^{(j)}(s)} \min_{1\le k\le n} \rho^2_{\dot{u}_k}, \quad s_j \le s \le s_f. \tag{10.151d}$$

(vii) **Minimum force and moment rate**

Corresponding to (10.146a) we may consider the integral

$$E\Big(J\big(\dot{u}(\cdot)\big|\mathfrak{A}_{t_j}\big) := E\Big(\int_{t_j}^{t_f} \big\|\dot{u}(t)\big\|^2 dt\Big|\mathfrak{A}_{t_j}\Big). \tag{10.152a}$$

Again with (10.139b) we find

$$\begin{aligned} E\Big(J\big(\dot{u}(\cdot)\big|\mathfrak{A}_{t_j}\big) &= E\Big(\int_{t_j}^{t_f} \Big\|\frac{\partial u_e}{\partial s}\big(p_D(\omega), s; q_e(\cdot), \beta(\cdot)\big)\sqrt{\beta(s)}\Big\|^2 \frac{ds}{\sqrt{\beta(s)}}\Big|\mathfrak{A}_{t_j}\Big) \\ &= \int_{s_j}^{s_f} E\Big(\Big\|\frac{\partial u_e}{\partial s}\big(p_D(\omega), s; q_e(\cdot), \beta(\cdot)\big)\Big\|^2\Big|\mathfrak{A}_{t_j}\Big)\sqrt{\beta(s)}ds \\ &= \int_{s_j}^{s_f} \Big(\mathrm{var}\Big(\frac{\partial u_e}{\partial s}\big(p_D(\omega), s; q_e(\cdot), \beta(\cdot)\big)\Big|\mathfrak{A}_{t_j}\Big) \\ &\qquad + \Big\|\frac{\partial u_e}{\partial s}\big(\overline{p}_D^{(j)}, s; q_e(\cdot), \beta(\cdot)\big)\Big\|^2\Big)\sqrt{\beta(s)}ds. \end{aligned} \tag{10.152b}$$

If we consider

$$E\Big(\tilde{J}\big(\dot{u}(\cdot)\big)\Big|\mathfrak{A}_{t_j}\Big) := E\Big(\int_{t_j}^{t_f} \big\|\dot{u}(t)\big\|dt\Big|\mathfrak{A}_{t_j}\Big), \tag{10.153a}$$

then we get

$$E\Big(\tilde{J}\big(\dot{u}(\cdot)\big)\Big|\mathfrak{A}_{t_j}\Big) = \int_{s_j}^{s_f} E\Big(\Big\|\frac{\partial u_e}{\partial s}\big(p_D(\omega), s; q_e(\cdot), \beta(\cdot)\big)\Big\|\Big|\mathfrak{A}_{t_j}\Big)ds. \tag{10.153b}$$

References

1. Arimoto, S.: Control Theory of Non-Linear Mechanical Systems. Clarendon Press, Oxford (1996)
2. Åström, K., Wittenmark, B.: Adaptive Control. Addison-Wesley (1995)
3. Aurnhammer, A., Marti, K.: Ostp/fortran-program for optimal stochastic trajectory planning of robots. Technical report, UniBw München (2000)

4. Bastian, G., et al.: Theory of Robot Control. Springer, Berlin (1996)
5. Bauer, H.: Wahrscheinlichkeitstheorie und Grundzüge der Masstheorie. Walter de Gruyter & Co., Berlin (1968)
6. Bernhardt, R., Albright, S.: Robot Calibration. Chapman and Hall, London (1993)
7. Bobrow, J.: Optimal robot plant planning using the minimum-time criterion. IEEE J. Robot. Autom. **4**(4), 443–450 (1988). https://doi.org/10.1109/56.811
8. Bobrow, J., Dubowsky, S., Gibson, J.: Time-optimal control of robotic manipulators along specified paths. Int. J. Robot. Res. **4**, 3–17 (1985)
9. Bohlin, R.: Motion Planning for Industrial Robots. Chalmers University, Goeteborg (1999)
10. Chen, Y.: Solving robot trajectory planning problems with uniform cubic b-splines. Optim. Control Appl. Methods **12**(4), 247–262 (1991). https://doi.org/10.1002/oca.4660120404
11. Craig, J.: Adaptive Control of Mechanical Manipulators. Addison-Wesley (1988)
12. Dullerud, G., Paganini, F.: A Course in Robust Control Theory. Springer, New York (2000)
13. Gessner, P., Spreman, K.: Optimierung in Funktionenräumen. Lecture Notes in Economics and Mathematical Systems, vol. 64. Springer, Berlin (1972)
14. Haubach-Lippmann, C.: Stochastische Strukturoptimierung flexibler Roboter, Fortschrittberichte VDI, Reihe 8, vol. 706. VDI Verlag, Düsseldorf (1998)
15. Holtgrewe, D.: Adaptive Regelung flexibler Roboter. Igel Verlag, Paderborn (1996))
16. Hoppen, P.: Autonome mobile Roboter. B.I. Wissenschaftsverlag, Mannheim (1992)
17. Hwang, Y.K., Ahuja, N.: Gross motion planning-a survey. ACM Comput. Surv. **24**(3), 219–291 (1992). https://doi.org/10.1145/136035.136037
18. Isermann, R.: Identifikation dynamischer Systeme. Springer, Berlin (1988)
19. Johanni, R.: Optimale Bahnplanung bei Industrierobotern, Fortschrittberichte VDI, Reihe 18, vol. 51. VDI-Verlag, Düsseldorf (1988)
20. Kall, P.: Stochastic Linear Programming. Springer, Berlin (1976)
21. Kall, P., Wallace, S.: Stochastic Programming. Stochastic Programming. Wiley, Chichester (1994)
22. Khalil, H.: Nonlinear Systems. Maxmillan Publishing Company, New York (1992)
23. Lin, K.: Zur adaptiven und langzeitprädikativen Regelung mit einem statistischen Optimierungskriterium. Ph.D. thesis, TU Hamburg-Harburg (1997)
24. Marti, K.: Approximationen stochastischer Optimierungsprobleme. Hain Konigstein/Ts (1979)
25. Marti, K.: Path planning for robots under stochastic uncertainty. Optimization **45**(1–4), 163–195 (1999). https://doi.org/10.1080/02331939908844432
26. Marti, K., Kall, P. (eds.): Stochastic Programming Methods and Technical Applications. Lecture Notes in Economics and Mathematical Systems, vol. 458. Springer, Berlin (1998)
27. Marti, K., Kall, P. (eds.): Stochastic Programming: Numerical Techniques and Engineering Applications. Lecture Notes in Economics and Mathematical Systems, vol. 423. Springer, Berlin (1998)
28. Marti, K., Qu, S.: Optimal trajectory planning for robots under the consideration of stochastic parameters and disturbances. J. Intell. Rob. Syst. **15**, 19–23 (1996)
29. Marti, K., Qu, S.: Path planning for robots by stochastic optimization methods. J. Intell. Rob. Syst. **22**, 117–127 (1998)
30. Marti, K., Qu, S.: Adaptive stochastic path planning for robots–real–time optimization by means of neural networks. In: M. Polis, et al. (eds.) Systems Modeling and Optimization, Proceed. 18th TC7 Conference, Research Notes in Mathematics, pp. 486–494. Chapman and Hall/CRC, Boca Raton (1999)
31. Miesbach, S.: Bahnführung von Robotern mit Neuronalen Netzen. Ph.D. thesis, TU München, Fakultät für Mathematik (1995)
32. Pai, D., Leu, M.: Uncertainty and compliance of robot manipulators with application to task feasibility. Int. J. Robot. Res. **10**(3), 200–212 (1991)
33. Pfeiffer, F., Johanni, R.: A concept for manipulator trajectory planning. IEEE J. Robot. Autom. **3**(2), 115–123 (1987)
34. Pfeiffer, F., Reithmeier, E.: Roboterdynamik. Teubner Stuttgart (1987)

35. Pfeiffer, R., Richter, K.: Optimal path planning including forces at the gripper. J. Intell. Rob. Syst. **3**, 251–258 (1990)
36. Qu, S.: Optimal Bahnplanung unter Berücksichtigung stochastischer Parameterschwankungen, Fortschrittberichte VDI, Reihe 8, vol. 472. VDI-Verlag, Düsseldorf (1995))
37. Qu, Z., Dawson, D.: Robust Tracking Control of Robot Manipulators. IEEE Press, New York (1996)
38. Redfern, D., Goh, C.: Feedback control of state constrained optimal control problems. In: Dolezal, J., Fidler, J. (eds.) System Modelling and Optimization, pp. 442–449. Chapman and Hall, London (1996)
39. Reif, K.: Steuerung von nichtlinearen Systemen mit Homotopie-Verfahren, Fortschrittberichte VDI, Reihe 8: Mess-, Steuerungs- und Regelungstechnik, vol. 631. VDI Verlag, Düsseldorf (1997)
40. Schilling, K., et al.: The European development of small planetary mobile vehicle. Space Technol. **3**(17), 151–162 (1997)
41. Schilling, K., Flury, W.: Autonomy and on-board mission management aspects for the cassini titan probe. Acta Astronaut. **21**(1), 55–68 (1990). https://doi.org/10.1016/0094-5765(90)90106-U
42. Schilling, R.: Fundamentals of Robotics. Analysis and Control. Prentice Hall, London (1990)
43. Schlöder, J.: Numerische Methoden zur Behandlung hochdimensionaler Aufgaben der Parameteridentifizierung. Ph.D. thesis, Universität Bonn, Math.-Naturwissenschaftliche Fakultät (1998)
44. Schröer, K.: Identifikation von Kalibrierungsparametern kinematischer Ketten. Carl Hanser, München Wien (1993)
45. Sciavicco, L., Siciliano, B.: Modeling and Control of Robot Manipulators. Springer, London (2000)
46. Shin, K., McKay, N.: Minimum-time control of robotic manipulators with geometric path constraints. IEEE Trans. Autom. Control **30**(6), 531–541 (1985). https://doi.org/10.1109/TAC.1985.1104009
47. Slotine, J.J., Li, W.: Applied Nonlinear Control. Prentice-Hall Int. Inc., Englewood Chiffs (1991)
48. Stengel, R.: Stochastic Optimal Control: Theory and Application. Wiley, New York (1986)
49. Stone, H.: Kinematic Modeling, Identification, and Control of Robotic Manipulators. Kluwer Academic Publishing, Boston (1987)
50. Weinmann, A.: Uncertain Models and Robust Control. Springer, New York (1991)
51. Weinzierl, K.: Konzepte zur Steuerung und Regelung nichtlinearer Systeme auf der Basis der Jacobi-Linearisierung. Verlag Shaker, Aachen (1995)

Chapter 11
Machine Learning Under Stochastic Uncertainty

Abstract New methods for machine learning under stochastic uncertainty, especially for regression problems under uncertainty are described in this chapter. Given a set of input-output data, a certain, often parametric set of functions is adapted by evaluating and then minimizing the approximation error by quadratic cost functions. Here, instead of quadratic cost functions, sublinear cost functions, involving, e.g., the maximum absolute error, are taken into account. In this case the regression problem under stochastic uncertainty yields a stochastic linear program with a dual decomposition data structure which enables the use of very efficient linear programming algorithms. Two and multi-group classification problems are considered in the second part of this chapter. Here, the separation of the data points in a certain space $\mathbb{R}^n$ and the representation of the groups or classes of data points is described by means of hyperplanes in $\mathbb{R}^n$. Instead of the often used discrete data points, for the classification process convex or stochastic convex hulls of the given data points are taken into account.

11.1 Foundations

In machine meaning [2] one has the problem to determine an unknown or partly known functional relation, cf. Fig. 11.1,

$$y = h(x) \tag{11.1a}$$

between a stochastic input n-vector x, called regressor in regression analysis and a stochastic output m-vector y, also called regressand.

For the estimation of the function $h : \mathbb{R}^n \to \mathbb{R}^m$, called response function in response surface methods (RSM) [12], besides eventually some a priori information on h, one has some data, as observations, measurements or scenarios of a discrete distribution, etc.,

$$(x_i, y_i),\ y_i = h(x_i) + \epsilon_i \text{ (error)}, \quad i = 1, 2, \ldots r, \tag{11.1b}$$

of the input-output pair $(x, y),\ y = h(x)$.

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9_11

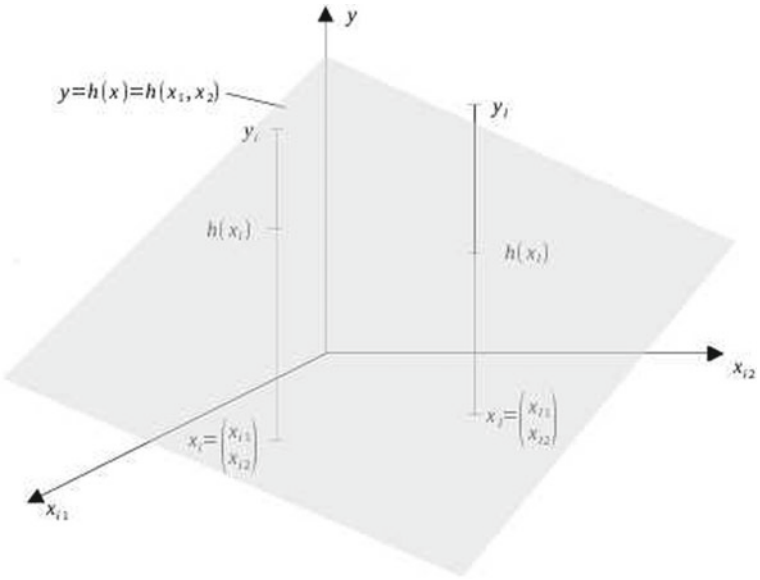


Fig. 11.1 Response function h, data (x_i, y_i)

For the estimation of the response function $y = h(x)$, usually parametric models

$$h = h(x; \beta) \tag{11.2a}$$

are used, which are linear in the parameter vector or matrix β, B and linear, quadratic or, more general, nonlinear in x:

$$h(x; \beta) = \begin{cases} \beta_0 + \beta_I^T x, \\ \beta_0 + \beta_I^T x + \frac{1}{2} x^T B x, \\ \beta_0 + \sum_{k=1}^{p} \beta_k \varphi_k(x) \end{cases} \tag{11.2b}$$

with parameters $\beta_0, \beta_k,\ k = 1, 2, \ldots, p$ and/or vectorial, matrix parameters $\beta_I = (\beta_1, \beta_2, \ldots, \beta_n)^T$, $B = (b_{k,l})_{1 \le k,l \le n}$ and certain, linear/nonlinear functions $\varphi_k = \varphi_k(x),\ k = 1, \ldots, p$ in x.

For a vectorial response function

$$y = y(x; \beta) = (y_1(x, \beta), y_2(x, \beta), \ldots, y_m(x, \beta))^T \tag{11.3}$$

with components as defined in (11.2b), having a joint parameter vector/matrix β, in the general case we may write

$$y = y(x; \beta_0, \beta_I) = \beta_0 + \varphi(x)\beta_I = (I, \varphi(x)) \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix}, \tag{11.4a}$$

where $\beta_0 = (\beta_{01}, \beta_{02}, \ldots, \beta_{0,m})^T$ is the initial parameter vector, $\beta_I = (\beta_1, \beta_2, \ldots, \beta_p)^T$, I is the unit matrix and $\varphi = \varphi(x)$ is the (m, p)-matrix

$$\varphi(x) = \begin{pmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \vdots \\ \varphi_m(x) \end{pmatrix} = \begin{pmatrix} \varphi_{11}(x) & \ldots & \varphi_{1p}(x) \\ \varphi_{21}(x) & \ldots & \varphi_{2p}(x) \\ \vdots & & \vdots \\ \varphi_{m1}(x) & \ldots & \varphi_{mp}(x) \end{pmatrix} \tag{11.4b}$$

with given functions $\varphi_{lk} = \varphi_{lk}(x)$, $l = 1, \ldots, m$, $k = 1, \ldots, p$. According to the stochastic data (11.1b), including measurements, observational and modeling errors, for the unknown or partly known parameter vectors β_0, β_I with (11.4a) and (11.4b) we have the following relations:

$$y_i \approx \beta_0 + \varphi(x_i)\beta_I, \; i = 1, 2, \ldots r. \tag{11.5a}$$

with the rn-, rm-vectors and the $(rm, m + p)$-matrix

$$X := \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix}, \qquad Y := \begin{pmatrix} y_1 \\ \vdots \\ y_r \end{pmatrix}, \qquad H(X) := \begin{pmatrix} I\varphi(x_1) \\ \vdots \\ I\varphi(x_r) \end{pmatrix}, \tag{11.5b}$$

the above relations (11.5a) can be represented by

$$Y \approx H(X) \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix}. \tag{11.5c}$$

Considering standard regression techniques, the deviation between the vector Y of all observed outputs y_i, $i = 1, \ldots, r$, and the vector $H(X) \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix}$ of the predicted outputs by the model (11.5a)–(11.5c) with respect to the vector X of all inputs x_i, $i = 1, \ldots, r$ is evaluated by means of a loss function $q = q(z)$, as, e.g., the squared Euclidean norm $q(z) = \|z\|_E = \sum_{i=1}^{r} \|z_i\|_E^2$.

This leads to the optimization problem

$$\min_{(\beta_0, \beta_I) \in D} q(Y - H(X) \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix}). \tag{11.5d}$$

11.2 Stochastic Optimization Methods in Machine Learning

The data $(x_i, y_i),\ i = 1, \ldots, r$, as stated in (11.1b), can be interpreted as realizations of a pair

$$(X, Y) = (X(\omega), Y(\omega)),\ \omega \in (\Omega, \mathfrak{A}, \mathcal{P}), \tag{11.6a}$$

of stochastic scalar or vectorial input-output variables on a probability space $(\Omega, \mathfrak{A}, \mathcal{P})$.

Taking possible constraints $\beta \in D$ and/or costs $c(\beta) = c(\beta_0, \beta_I)$ into account for the parameter vector

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix} \tag{11.6b}$$

with a loss function $q = q(z)$, the machine learning or regression problem can be formulated by the stochastic optimization problem

$$\min_{\beta \in D} c(\beta) + Eq\left(Y(\omega) - H(X(\omega)\beta\right), \tag{11.6c}$$

where

$$H(X(\omega))\beta = \beta_0 + \varphi(X(\omega))\beta_I, \tag{11.6d}$$

E denotes the expectation operator and $X = X(\omega) \in \mathbb{R}^n$, $Y = Y(\omega) \in \mathbb{R}^m$ denotes the underlying input-output pair.

11.2.1 Least Squares Estimation of the Parameter Vector

One of the most frequently used loss functions is the squared Euclidean norm $q(z) = \|z\|_E^2 = z^T z$. Assuming zero parameter costs $c(\beta) = 0$ and no constraints for $\beta = (\beta_0, \beta_I)^T$, the estimation problem (11.6c), (11.6d) reads

$$\min_{\beta_0, \beta_I} E \left\| Y(\omega) - H(X(\omega)) \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix} \right\|_E^2. \tag{11.7a}$$

Due to the strict convexity of the loss function $q(z) = \|z\|_E^2$, the necessary and sufficient optimality condition reads

$$E\,H(X(\omega))^T H(X(\omega)) \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix} = E\,H(X(\omega))^T Y(\omega).$$

This yields the following result:

Lemma 11.1 *If the loss function q is defined by the squared Euclidean norm and the matrix $E\,H(X(\omega))^T H(X(\omega))$ is regular, then the optimal parameter vector is given by*

$$\begin{pmatrix} \beta_0^* \\ \beta_I^* \end{pmatrix} = \left(E\,H(X(\omega))^T H(X(\omega))\right)^{-1} E\,H(X(\omega))^T Y(\omega). \tag{11.7b}$$

11.3 Estimation with Sublinear Loss Function $q = q(z)$

In addition to the stochastic optimization problem (11.6c)–(11.6d), we now consider the parameter optimization problem in the following form, where $Z = Z(\omega) \in \mathbb{R}^m$, $\omega \in (\Omega, \mathfrak{A}, \mathcal{P})$, denotes the approximated and realized output

$$\min\ c(\beta_0, \beta_I) + Eq(Z(\omega)) \tag{11.8a}$$

$$\text{s.t. } H(X(\omega)) \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix} + Z(\omega) = Y(\omega) \tag{11.8b}$$

$$(\beta_0, \beta_I)^T \in D. \tag{11.8c}$$

Here, $q = q(z)$ is a sublinear function [7, 9] having the following properties:

$$\begin{aligned} q(\lambda z) &= \lambda q(z), & \lambda \geq 0,\ z \in \mathbb{R}^m \\ q(z + w) &\leq q(z) + q(w), & z, w \in \mathbb{R}^m. \end{aligned}$$

Sublinear functionals can be represented in the following way [8]:

(a) *Maximum (supremum) of linear functions*
Let c_j, $j \in J$ be a finite or infinite number of vectors $c_j \in \mathbb{R}^m$. Then

$$q(z) := \max_{j \in J} (\sup) c_j^T z,\ z \in \mathbb{R}^m \tag{11.9a}$$

is a sublinear function on $\mathbb{R}^m$, assuming that for an infinite index set the supremum is finite.

(b) *Distance or Minkowski functional*
For a given closed, convex set $K \in \mathbb{R}^m$ containing the origin as an interior point, the distance or Minkowski functional $q(z) = q_K(z)$ is defined by

$$q(z) := \inf\{\lambda > 0 : \frac{z}{\lambda} \in K\},\ z \in \mathbb{R}^m. \tag{11.9b}$$

(c) *Sublinear functions generated by linear programs*
Let M an (m, s)-matrix such that

$$\left\{w \in \mathbb{R}^m \,:\, Mw = z,\ w \geq 0\right\} \neq \emptyset \text{ for all } z \in \mathbb{R}^m \tag{11.9c}$$

and $\gamma \geq 0,\ \gamma \in \mathbb{R}^s$. Then

$$q(z) := \min\left\{\gamma^T w \,:\, Mw = z,\ w \geq 0\right\},\ z \in \mathbb{R}^m \tag{11.9d}$$

i.e., the solution of the linear program

$$\begin{array}{ll} \min & \gamma^T w \\ \text{s.t.} & Mw = z \\ & w \geq 0 \end{array} \tag{11.9e}$$

is a sublinear function on $\mathbb{R}^m$.

Remark 11.1 For the interrelation of the three representations of sublinear functions, see [7, 8].

Some sublinear functions $q = q(z)$, as the maximum norm, also called Chebyshev norm, i.e., $q(z) = \|z\|_\infty = \sup |z_j|$, evaluate the components $z_j,\ j \in J$ of a deviation vector $z = (z_j)_{j \in J}$ in a *uniform way*.

Remark 11.2 The matrix M is called *recourse* matrix in cases where (11.9c)–(11.9d) models a correction step, see later.

Using the LP version (11.9d) for the representation of the sublinear function, the estimation problem (11.8a)–(11.8c) can be represented, using a possibly random matrix $M = M(\omega)$ and a random cost vector $\gamma = \gamma(\omega)$, by

$$\min \quad c(\beta_0, \beta_I) + E\gamma(\omega)^T w(\omega) \tag{11.10a}$$

$$\text{s.t.} \quad H(X(\omega))\begin{pmatrix}\beta_0 \\ \beta_I\end{pmatrix} + M(\omega)w(\omega) = Y(\omega) \qquad \text{a.s.} \tag{11.10b}$$

$$\begin{pmatrix}\beta_0 \\ \beta_I\end{pmatrix} \in D,\ w(\omega) \geq 0, \qquad \text{a.s.} \tag{11.10c}$$

11.3.1 Representation by a Stochastic Linear Optimization Problem (SLOP)

Under weak assumptions (11.10a)–(11.10c) can be represented by an (LP):

- Suppose that the cost function $c = c(\beta_0, \beta_I)$ is defined by the sublinear function, c.f. (11.9a),

$$c(\beta_0, \beta_I) := \max_{1 \le l \le L} c_l^T \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix}, \tag{11.11a}$$

where $c_l,\ l = 1, \ldots, L$, are given $(m + p)$-vectors.

- Moreover, assume $(\beta_0, \beta_I)^T \in D$ is given by

$$G \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix} \le g. \tag{11.11b}$$

- Defining the $(L, m + p)$-matrix C and the L-vector e_L by

$$C := \begin{pmatrix} c_1^T \\ \vdots \\ c_L^T \end{pmatrix}, \qquad e_L := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \tag{11.11c}$$

and selecting auxiliary vectorial variables ϵ, η,

for problem (11.8a)–(11.8c) we have the following representation [9]

Lemma 11.2 *Using* (11.11a)–(11.11c), *problem* (11.8a)–(11.8c) *can be represented in the form of a stochastic linear optimization problem (SLOP)*

$$\min \quad c + E\gamma(\omega)^T w(\omega) \tag{11.12a}$$

$$s.t. \quad H(X(\omega)) \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix} + M(\omega) w(\omega) = Y(\omega) \qquad a.s. \tag{11.12b}$$

$$C \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix} - c e_L + \delta = 0 \tag{11.12c}$$

$$G \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix} + \eta = g \tag{11.12d}$$

$$\delta, \eta \ge 0, \quad w(\omega) \ge 0 \qquad a.s. \tag{11.12e}$$

Remark 11.3 (*SLOP*) Condition (11.12a) and (11.12c) mean that the objective function of the SLOP (11.12a)–(11.12e),

$$f \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix} := \max_{1 \le l \le L} c_l^T \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix} + E\gamma(\omega)^T w(\omega), \tag{11.13}$$

i.e., the sum of the costs (11.11a) for the parameters $\beta = (\beta_0, \beta_I)^T$ and the expected costs (11.9d) for the deviation between the actual realization $Y(\omega)$ of the output Y and the approximated output $H(Y(\omega)) \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix}$ have to be minimized.

11.3.2 Numerical Solution of the (SLOP)

A main procedure for the numerical solution of problems of the type (11.12a)–(11.12e) is based on the discretization of the probability measure $\mathcal{P}$, see, e.g., [10, Chap. 2].

Here, $\mathcal{P}$ is approximated by discrete distributions

$$\mathcal{P}_d := \sum_{j=1}^{r} \alpha_j \epsilon_{\omega_j}, \qquad \sum_{j=1}^{r} \alpha_j = 1, \qquad \alpha_j \geq 0, \qquad 1 \leq j \leq r, \tag{11.14a}$$

where ϵ_{ω_j} denotes the one-point measure at point ω_j, and $\alpha_j = \mathcal{P}(\Omega_j)$ are the probabilities of a certain partition $\bigcup_{j=1}^{r} \Omega_j = \Omega$ of Ω with $w_j \in \Omega_j,\ j = 1, \ldots, r$.

Defining then $x_j, y_j, w_j, M_j,\ j = 1, \ldots, r$ by

$$\begin{aligned} x_j &:= X(\omega_j), & y_j &:= Y(\omega_j), & w_j &:= w(\omega_j), \\ M_j &:= M_j(\omega_j), & \gamma_j &:= \gamma(\omega_j), \end{aligned} \tag{11.14b}$$

problem (11.12a)–(11.12e) can be approximated by

$$\min \quad c + \sum_{j=1}^{r} x_j \gamma_j^T w_j \tag{11.15a}$$

$$\text{s.t.} \quad H(x_j)\beta + M_j w_j = y_j, \quad j = 1, \ldots, r \tag{11.15b}$$

$$c\beta - ce_L + \delta = 0 \tag{11.15c}$$

$$G\beta + \eta = g \tag{11.15d}$$

$$w_j \geq 0,\ j = 1, \ldots r, \quad \delta \geq 0, \eta \geq 0, \tag{11.15e}$$

with $\beta := \begin{pmatrix} \beta_0 \\ \beta_I \end{pmatrix}$.

Next, basic properties of problem (11.15a)–(11.15e) are given:

Theorem 11.1 (Stochastic Linear Program (SLP))

(a) The discretized problem (11.15a)–(11.15e) is a linear program (LP).

(b) Increasing the refinement of the discretization of the probability space $(\Omega, \mathfrak{A}, \mathcal{P})$, the optimal solution β_d^* and the related optimal value of (11.15a)–(11.15e) converge to the optimal solution, the optimal value, resp., of (11.12a)–(11.12e).

Proof The first assertion can be seen directly in (11.15a)–(11.15e). The second one can be found in the literature on stochastic linear programming (SLP), see, e.g., [4, 5]. □

Table 11.1 Data structure of the (SLP) (11.15a)–(11.15e)

β	c	δ	η	w_1	w_2	w_3	...	w_r	1
0	1	0	0	$\alpha_1\gamma_1$	$\alpha_2\gamma_2$	$\alpha_3\gamma_3$	...	$\alpha_r\gamma_r$	0
$H(x_1)$	0	0	0	M_1					y_1
$H(x_2)$	0	0	0		M_2				y_2
$H(x_3)$	0	0	0			M_3			y_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$				$\ddots$		$\vdots$
$H(x_r)$	0	0	0					M_r	y_r
C	$-e_L$	I	0						0
G	0	0	I						g

Another important property for the solution (11.15a)–(11.15e) follows from its special data structure, see Table 11.1 The following variables and data occur:

(i) In the first line all variables $\beta, c, \delta, \eta, w_1, \ldots, w_r$ are shown.
(ii) In the second line the coefficients of the objective function appear.
(iii) In the following large block the discretized input-output equation involving the unknown parameters to be determined occurs.
(iv) In the next to last line the coefficients for the representation of the objective function C as a maximum of linear functions of β are shown.
(v) The last line shows the coefficients of possible inequality constraints of the parameter vector β.

According to Table 11.1, (SLP) (11.15a)–(11.15e) has the following data structure which guarantees considerable advantages in the numerical solution of this LP, see [5, 11].

Theorem 11.2 *The LP* (11.15a)–(11.15e) *has a dual decomposition or dual block angular data structure.*

Proof See the consideration of (SLOP), e.g., in [4, 5, 11]. □

11.3.3 Two-Stage Stochastic Linear Programs (SLP)

In problem (11.12a)–(11.12e), (11.15a)–(11.15e), the system of linear equations (11.12b), (11.15b), resp. , together with the second cost term in (11.12a), (11.15a), can be interpreted as a cost-based evaluation of the accuracy of the approximation of the output $Y(\omega)$ by $H(X(\omega))\beta$, $\omega \in \Omega$, y_i by $H(x_i)\beta$, $i = 1, \ldots, r$, respectively.

Assume now, that after taking the coefficient vector β at an initial time t_0, the random data input-output vector $(X(\omega), Y(\omega))$ is realized at a later time t_1. This enables then an improvement of the initial function approximation

$$h(x) \approx h(x;\beta) = H(x)\beta = \beta_0 + \varphi(x)\beta_I, \tag{11.16a}$$

see (11.1a)–(11.1b) and (11.2a)–(11.2b), (11.3), (11.4a)–(11.4b), by adding the extended correction term

$$\tilde{M}\tilde{w} := \psi(x)b + Mw = \sum_{l=1}^{L} \psi(x)\beta_l + Mw, \tag{11.16b}$$

where the first term is used to improve the analytical approximation of the input-output function $y = h(x)$, hence

$$\hat{h}(x) = \varphi(x)\beta \to \hat{h} = \varphi(x)\beta + \psi(x)b, \tag{11.16c}$$

adding, e.g., quadratic terms to an initially linear approximation of h. Moreover, the second term in (11.16b) is devoted to measure, minimize, resp. , the remaining analytical and measurement/observational errors.

Using the extended correction term (11.16b) and taking, for simplification, linear cost functions for the evaluation of the coefficient vectors β and b, corresponding to (11.12a)–(11.12e) the (SLOP) reads

$$\begin{aligned}
\min_{\beta_+,\beta_-,b_+,b_-,w} \quad & c_+^T\beta_+ + c_-^T\beta_- + E\left(d_+^T(\omega)b_+(\omega) + d_-^T(\omega)b_-(\omega)\right. \\
& \left. + \gamma(\omega)^T w(\omega)\right) && \text{(11.17a)} \\
\text{s.t.} \quad & H(X(\omega))(\beta_+ - \beta_-) + \psi(X(\omega))(b_+(\omega) - b_-(\omega)) \\
& + M(\omega)w(\omega) = Y(\omega) \quad \text{a.s.} && \text{(11.17b)} \\
& G(\beta_+ - \beta_-) + \eta = g && \text{(11.17c)} \\
& \beta_+, \beta_-, \eta \geq 0, \quad b_+(\omega), b_-(\omega), w(\omega) \geq 0 \quad \text{a.s.} && \text{(11.17d)}
\end{aligned}$$

Remark 11.4 The vectors β, $b(\omega)$ of coefficients in the approximation of the input-output function $y = h(x)$, sf. (11.1a)–(11.1b) are represented by

$$\beta = \beta_+ - \beta_-, \quad b(\omega) = b_+(\omega) - b_-(\omega) \tag{11.18a}$$

with nonnegative vectors $\beta_+, \beta_-, b_+(\omega), b_-(\omega) \geq 0$. Note, that in the present 2-stage setting of the (SLOP) the vectors $b(\omega)$, $b_+(\omega)$ and $b_-(\omega)$ may be random. Moreover, c_+, c_-, $d_+ = d_+(\omega), d_- = d_-(\omega)$, resp., are deterministic, stochastic, resp., cost vectors.

Approximating, corresponding to (11.12a)–(11.12e), the 2-stage (SLOP) (11.17a)–(11.17d) by means of discretization of the probability distribution $\mathcal{P}$, see (11.14a)–(11.14b), we set

$$\begin{aligned}
& b_{+j} := b_+(\omega_j), \quad b_{-j} := b_-(\omega_j), \quad M_j := M(\omega_j), \quad w_j := w(\omega_j), \\
& d_{+j} := d_+(\omega_j), \quad d_{-j} := d_-(\omega_j), \quad \gamma_j := \gamma(\omega_j), \qquad j = 1,\ldots,r.
\end{aligned} \tag{11.18b}$$

Corresponding to (SLP) (11.15a)–(11.15e) the 2-stage (SLOP) (11.17a)–(11.17d) is approximated, cf. (11.14a)–(11.14b), by discretization of the probability distribution $\mathcal{P}$, by the 2-stage stochastic linear program

$$\min_{\substack{\beta_+,\beta_-,\eta \\ b_{+j},b_{-j},w_j}} \quad c_+^T\beta_+ + c_-^T\beta_- + \sum_{j=1}^{r} \alpha_j \left(d_{+j}^T b_{+j} + d_{-j}^T b_{-j} + \gamma_j^T w_j\right) \tag{11.19a}$$

$$\text{s.t.} \quad H(\alpha_j)(\beta_+ - \beta_-) + \psi(x_j)(b_{+j} - b_{-j}) + M_j w_j = y_j \; j = 1, \ldots r \tag{11.19b}$$

$$G(\beta_+ - \beta_-) + \eta = g \tag{11.19c}$$

$$\beta_+, \beta_-, \eta \geq 0, \quad b_{+j}, b_{-j}, w_j \geq 0, \quad j = 1, \ldots r \tag{11.19d}$$

Since the (SLP) (11.19a)–(11.19d) and (11.15a)–(11.15e) have the same LP structure, all properties which are valid for (11.15a)–(11.15e), see Theorem 11.2, hold for (11.19a)–(11.19d), too.

Especially, we mention the following property which is important for the numerical solution of (11.19a)–(11.19d).

Theorem 11.3 *The LP* (11.19a)–(11.19d) *has a dual decomposition or dual block angular data structure.*

Proof See, e.g., the consideration of the stochastic linear programs in [4, 5, 11]. □

11.4 Two and Multiple Group Classification Under Stochastic Uncertainty

Consider a set of data points x_i,

$$M = \{x_i \; : \; i = 1, \ldots, I\} \subset \mathbb{R}^n, \tag{11.20a}$$

which are realizations of a random n-vector $\zeta = \zeta(\omega)$ on a probability space $(\Omega, \mathfrak{A}, P)$. Assume that a certain grouping, cf. Fig. 11.2,

$$M = \bigcup_{j=1}^{J} M_j, \quad M_J := \left\{x_i : i \in I_j\right\}, \; j = 1, \cdots J \tag{11.20b}$$

into disjoint sets of points M_j, $j = 1, \ldots, J$, can be observed, where also the further realizations belong to one of the observed classes.

The problem now is to give a suitable mathematical description of the observed classification such that the class of a new data point can be predicted.

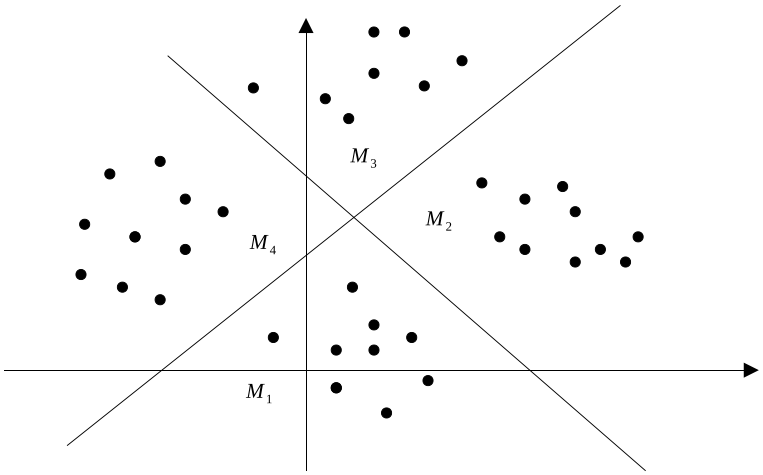


Fig. 11.2 Observed grouping of the data set M

As indicated in Fig. 11.2, the observed classes of data points can be represented in many cases [1–3, 6] by means of a certain number of straight lines (in $\mathbb{R}^2$) or hyperplanes in $\mathbb{R}^n$:

$$H_v := \left\{x \in \mathbb{R}^n \,:\, w^T x = b\right\}, \quad v = \begin{pmatrix} w \\ b \end{pmatrix} \in \mathbb{R}^{n+1}, \tag{11.21a}$$

$$v = v_l = \begin{pmatrix} w \\ b_l \end{pmatrix}, \quad l = 1, \ldots, L. \tag{11.21b}$$

The parameter vectors v_l, $l = 1, \ldots L$ are chosen in a way that point sets, representing the J classes, can be subdivided by the hyperplanes and their sides in a certain optimal sense, described as follows:

11.4.1 Two Classes ($J = 2, L = 1$)

Separating in case of two classes, cf. Fig. 11.3, the corresponding data sets M_1, M_2 with a hyperplane H_v, $v = \begin{pmatrix} w \\ b \end{pmatrix}$, for the parameter $(n+1)$-vector v, we have the basic condition

$$w^T x - b \geq 0, \quad x \in M_1, \tag{11.22a}$$

$$w^T x - b \leq 0, \quad x \in M_2. \tag{11.22b}$$

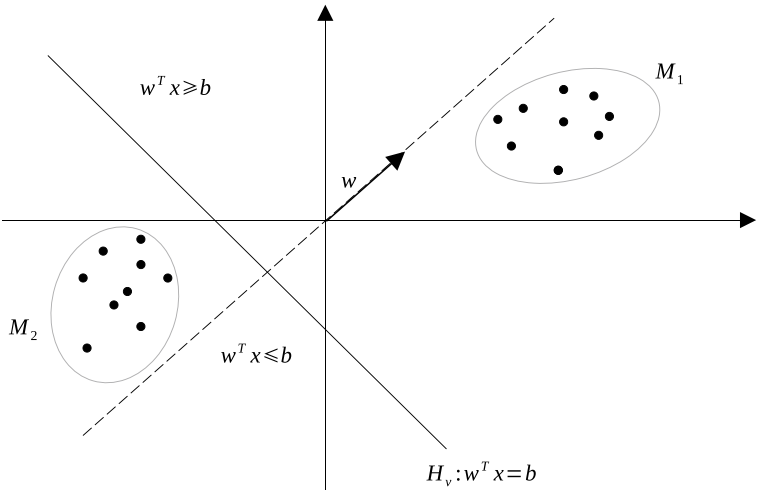


Fig. 11.3 Separation of data sets M_1, M_2, related to two classes, by the hyperplane H_v

Obviously, there are many possible separating hyperplanes H_v in general. Hence, in order to get a unique, sharp separation of the data sets M_1, M_2, we consider the minimum distance d_1^*, d_2^*, resp., between a separating hyperplane H_v and a point x of M_1, M_2, respectively.

The distances $d_1 = d_1(x),\ d_2 = d_2(\tilde{x})$ between a point $x \in M_1,\ \tilde{x} \in M_2$, resp., and the hyperplane H_v are given by, cf. Fig. 11.4

$$d_1(x) = \|x - u\| = \frac{w^T x - b}{\|x\|}, \quad x \in M_1, \tag{11.23a}$$

$$d_2(X) = \|\tilde{x} - \tilde{u}\| = \frac{b - w^T \tilde{x}}{\|w\|}, \quad \tilde{x} \in M_2, \tag{11.23b}$$

where $u, \tilde{u}$, resp., denotes the projection of a point $x \in M_1,\ \tilde{x} \in M_2$, resp., onto H_v. The minimum distances $d_1^*,\ d_2^*$, resp., between M_1, M_2, resp., and a separating hyperplane H_v then read

$$d_1^* = d_1^*(M_1) = \min_{x \in M_1} \frac{w^T x - b}{\|x\|}, \tag{11.24a}$$

$$d_2^* = d_2^*(M_2) = \min_{\tilde{x} \in M_2} \frac{b - w^T \tilde{x}}{\|\tilde{x}\|}. \tag{11.24b}$$

Consequently, the optimal separating hyperplane $H^* = H_{v^*}$ is achieved for equal and maximum distances d_1^*, d_2^*. Hence, we have the following conditions:

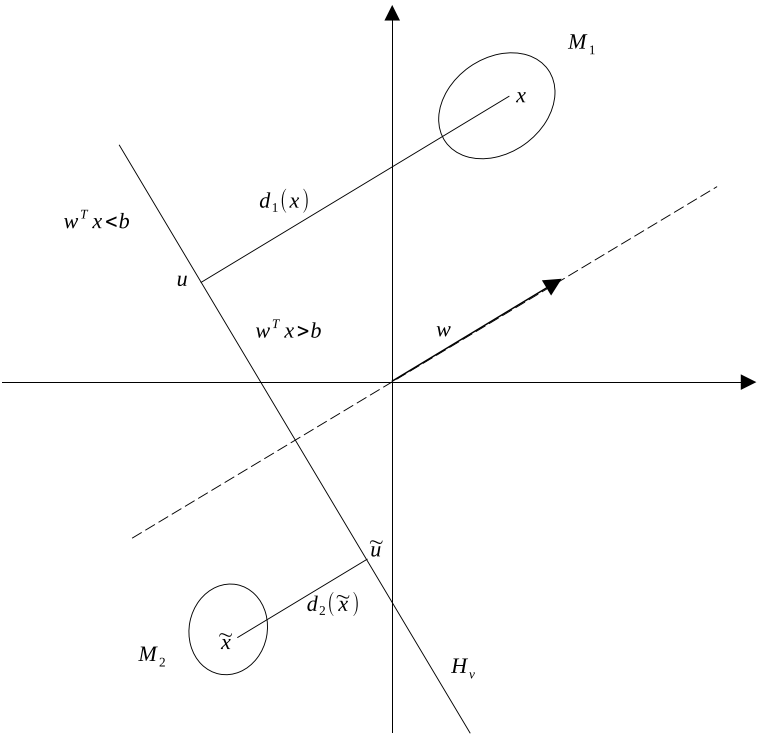


Fig. 11.4 Distances between a separating hyperplane H_v and a point x in M_1 or M_2

Theorem 11.4 (Optimal separating hyperplane $H^* = H_{v^*}$) *The parameter vector* $v = \begin{pmatrix} w^* \\ b^* \end{pmatrix}$ *of the optimal separating hyperplane* $H^* = H_{v^*}$ *is an optimal solution of the optimization problem*

$$\text{maximize} \quad \min_{x \in M_1} \frac{w^T x - b}{\|w\|} \tag{11.25a}$$

s.t.

$$w^T x - b \geq 0, \quad x \in M_1, \tag{11.25b}$$

$$w^T x - b \leq 0, \quad x \in M_2, \tag{11.25c}$$

$$\min_{x \in M_1} \frac{w^T x - b}{\|w\|} = \min_{x \in M_2} \frac{b - w^T x}{\|w\|}, \tag{11.25d}$$

$$\begin{pmatrix} w \\ b \end{pmatrix} \in \mathbb{R}^{n+1}. \tag{11.25e}$$

Remark 11.5 Because of Eqs. (11.24a), (11.24b), the equality condition $d_1^*(M_1) = d_2^*(M_2)$ is stated in the form (11.25d). Of course, it can also be given by

$$\min_{x \in M_1} w^T x - b = \min_{x \in M_2} b - w^T x. \tag{11.25f}$$

Corollary 11.1 *Simplified versions of the optimization problem* (11.25a)–(11.25f) *can be obtained if the variable b is replaced*

$$b := b_0 \tag{11.25g}$$

by a preselected fixed parameter value $b_0 \in \mathbb{R}$.

In the following we denote, cf. Fig. 11.5, by

$$x_1^* = x_1^*(w, b) \in M_1, \tag{11.26a}$$
$$x_2^* = x_2^*(w, b) \in M_2, \tag{11.26b}$$

an optimal solution (observed data point in M_1, M_2, resp.) of the internal optimization problems in condition (11.25d) or (11.25f).

Due to the uncertainties of stochastic variations of the observed data sets $M_1 = \{x_i : i \in I_1\}$, $M_2 = \{x_i : i \in I_2\}$, see (11.20a), (11.20b), we may replace (approximate) M_1, M_2, resp., by their convex hull

$$M_1^{\text{conv}} = \text{conv}(M_1), \qquad M_2^{\text{conv}} = \text{conv}(M_2) \tag{11.26c}$$

In this case the equations, see Fig. 11.5,

$$w^T x = b_1^* := w^T x_1^*, \tag{11.27a}$$
$$w^T x = b_2^* := w^T x_2^* \tag{11.27b}$$

denote the tangent hyperplanes to M_1^{conv}, M_2^{conv}, resp., at x_1^*, x_2^*, resp., parallel to the hyperplane H_v, $v = \begin{pmatrix} w \\ b \end{pmatrix}$ lying in th middle of the area between the two hyperplanes (11.27a), (11.27b), cf. Fig. 11.5.

According to the construction, with the hyperplanes (11.27a), (11.27b) we now obtain the following classification method for two class problems.

Theorem 11.5 (Classification rule for two class problems) *Let* $v^* = \begin{pmatrix} w^* \\ b^* \end{pmatrix}$ *be an optimal solution of the optimization problem* (11.25a)–(11.25f) *and* $x_1^* = x_1^*(v^*)$, $x_2^* = x_2^*(v^*)$ *the optimal solutions resulting from* (11.25d) *or* (11.25f).

To classify a new data point x, we may proceed as follows:

$$(I)\ \textit{If} \quad w^{*T} x > b_1^* := w^{*T} x_1^*(v^*), \quad \textit{then} \quad x \textit{ belongs to Class 1.} \tag{11.28a}$$
$$(II)\ \textit{If} \quad w^{*T} x < b_2^* := w^{*T} x_2^*(v^*), \quad \textit{then} \quad x \textit{ belongs to Class 2.} \tag{11.28b}$$

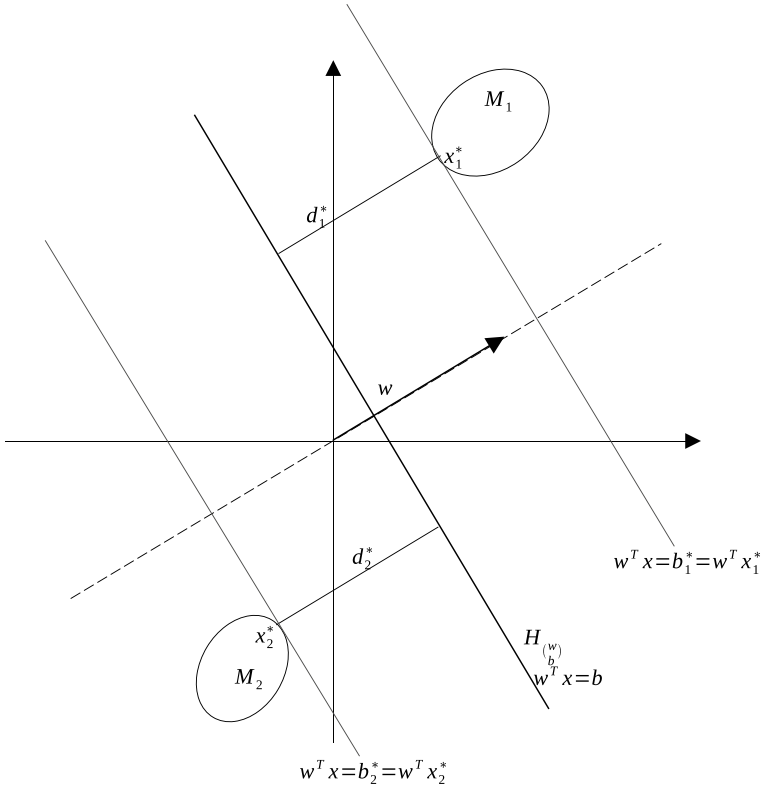


Fig. 11.5 Equal minimum distances $d_1^*(M_1) = d_2^*(M_2)$ between $H_{\binom{w}{b}}$ and M_1, M_2

11.5 Multi-classification

In case of data sets $M = \bigcup_{j=1}^{J} M_j$, cf. (11.20a), (11.20b), containing elements of a larger number, $J > 2$, of different properties, the structure of data set M, especially the configuration of the appearing data classes, plays an important role for the design of appropriate classifiers.

An important basic data structure depends on the property that the data set $M_j \subset \mathbb{R}^n$ of each class $j,\ j = 1, \ldots, J$, can be separated by a hyperplane H_j from the data points of the other classes $l \neq j$.

11.5.1 Reduction of a Multi-classifier to Several Two-Class Classifiers

As indicated by Fig. 11.6 , in certain cases for each class $j = 1, \ldots, J$, the data set M can be separated into the two subsets

$$M_j \qquad \text{and} \qquad \bigcup_{l \neq j} M_j, \quad j = 1, \ldots, J \tag{11.29a}$$

by a hyperplane. This allows the application of the described two-class classifier to the reduced two-class problem:

$$\text{class } j \qquad \text{and} \qquad \text{not class } j. \tag{11.29b}$$

Thus, we have the following result.

Theorem 11.6 (Classification rule for the reduction method of multi-classification problems to two-class classifiers) *Suppose that the data set M of a multi-classification problem has the structure described in Sect. 11.5.1. Then, for a given new data point x, by the J-th two-class classification, the class of x is known.*

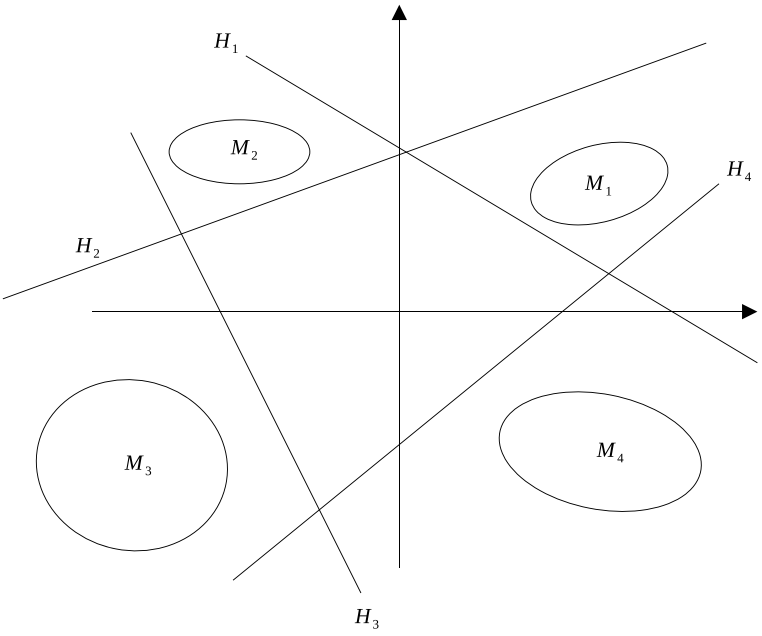


Fig. 11.6 Reduction of a four-class problem to four two-class problems

References

1. Bartkutė-Norkūnienė, V.: Stochastic optimization algorithms for support vector machines classification. Informatica **20**(2), 173–186 (2009). https://doi.org/10.15388/Informatica.2009.244
2. Bousquet, O., et al. (eds.): Advanced Lectures on Machine Learning. Springer Berlin (2004). https://doi.org/10.1007/b100712
3. Van den Burg, G., Groenen, P.: Gensvm: a generalized multiclass support vector machine. J. Mach. Learn. Res. **17**(224), 1–42 (2016). http://jmlr.org/papers/v17/14-526.html
4. Kall, P.: Stochastic Linear Programming. Springer, Berlin (1976)
5. Kall, P., Wallace, S.: Stochastic Programming. Stochastic Programming. Wiley, Chichester (1994)
6. Ma, Y., Guo, G. (eds.): Support Vector Machines Applications. Springer, Cham (2014)
7. Marti, K.: Entscheidungsprobleme mit linearem Aktionen- und Ergebnisraum. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **23**, 133–147 (1972)
8. Marti, K.: Approximationen der Entscheidungsprobleme mit linearer Ergebnisfunktion und positiv homogener, subadditiver Verlustfunktion. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **31**, 203–233 (1975)
9. Marti, K.: Approximationen stochastischer Optimierungsprobleme. Hain Konigstein/Ts (1979)
10. Marti, K.: Computation of efficient solutions of discretely distributed stochastic optimization problems. ZOR Methods Model. Oper. Res. **36**(3), 259–294 (1992). https://doi.org/10.1007/BF01415892
11. Mayer, J.: Stochastic Linear Programming Algorithms: A Comparison Based on a Model Management System. Gordon and Breach Science Publishers, Amsterdam (1998)
12. Myers, R.: Response Surface Methodology. Allyn and Bacon, Boston (1971)

Chapter 12
Stochastic Structural Optimization with Quadratic Loss Functions

Abstract Structural Analysis and Optimal Structural Design under Stochastic Uncertainty using Quadratic Cost Functions are treated in this chapter: Problems from plastic analysis and optimal plastic design are based on the convex, linear or linearized yield/strength condition and the linear equilibrium equation for the stress (state) vector. In practice one has to take into account stochastic variations of the vector $a = a(\omega)$ of model parameters (e.g., yield stresses, plastic capacities, external load factors, cost factors, etc.). Hence, in order to get robust optimal load factors x, robust optimal designs x, resp., the basic plastic analysis or optimal plastic design problem with random parameters has to be replaced by an appropriate deterministic substitute problem. As a basic tool in the analysis and optimal design of mechanical structures under uncertainty, a *state function* $s^* = s^*(a, x)$ of the underlying structure is introduced. The survival of the structure can be described then by the condition $s^* \leq 0$. Interpreting the state function s^* as a cost function, several relations s^* to other cost functions, especially quadratic cost functions, are derived. Bounds for the probability of survival p_s are obtained then by means of the Tschebyscheff inequality. In order to obtain robust optimal decisions x^*, i.e., maximum load factors, optimal designs insensitive with respect to variations of the model parameters $a = a(\omega)$, a direct approach is presented then based on the primary costs (weight, volume, costs of construction, costs for missing carrying capacity, etc.) and the recourse costs (e.g., costs for repair, compensation for weakness within the structure, damage, failure, etc.), where the above-mentioned quadratic cost criterion is used. The minimum recourse costs can be determined then by solving an optimization problem having a quadratic objective function and linear constraints. For each vector $a = a(\omega)$ of model parameters and each design vector x one obtains then an explicit representation of the best internal load distribution F^*. Moreover, also the expected recourse costs can be determined explicitly. The expected recourse function may be represented by means of a "generalized stiffness matrix". Hence, corresponding to an elastic approach, the expected recourse function can be interpreted here as a generalized expected compliance function, which depends on a generalized "stiffness matrix". Based on the minimization of the generalized compliance or the minimization of the expected total primary and recourse costs, explicit finite-dimensional parameter optimization problems are achieved for finding robust optimal design x^* or a maximal load factor x^*. The analytical properties of the resulting programming problem are

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9_12

discussed, and applications, such as limit load/shakedown analysis, are considered. Furthermore, based on the expected "compliance function", explicit upper and lower bounds for the probability p_s of survival.

12.1 Introduction

Problems from plastic analysis and optimal plastic design are based on the convex, linear, or linearized yield/strength condition and the linear equilibrium equation for the stress (state) vector. In practice one has to take into account stochastic variations of several model parameters. Hence, in order to get robust optimal decisions, the structural optimization problem with random parameters must be replaced by an appropriate deterministic substitute problem. A direct approach is proposed based on the primary costs (weight, volume, costs of construction, costs for missing carrying capacity, etc.) and the recourse costs (e.g., costs for repair, compensation for weakness within the structure, damage, failure, etc.). Based on the mechanical survival conditions of plasticity theory, a quadratic error/loss criterion is developed. The minimum recourse costs can be determined then by solving an optimization problem having a quadratic objective function and linear constraints. For each vector $a(\cdot)$ of model parameters and each design vector x, one obtains then an explicit representation of the "best" internal load distribution F^*. Moreover, also the expected recourse costs can be determined explicitly. It turns out that this function plays the role of a generalized expected *compliance function* involving a *generalized stiffness matrix*. For the solution of the resulting deterministic substitute problems, i.e., the minimization of the expected primary cost (e.g., volume, weight) subject to expected recourse cost constraints or the minimization of the expected total primary and recourse costs, explicit finite-dimensional parameter optimization problems are obtained. Furthermore, based on the quadratic cost approach, lower and upper bounds for the probability of survival can be derived.

In optimal plastic design of mechanical structure [2] one has to minimize a weight, volume or more general cost function c, while in limit load analysis [5] of plastic mechanical structures the problem is to maximize the load factor μ—in both cases—subject to the survival or safety conditions, consisting of the equilibrium equation and the so-called yield (feasibility) condition of the structure.

Thus, the objective function G_0 to be minimized is defined by

$$G_0(x) = \sum_{i=1}^{B} \gamma_{i0}(\omega) L_i A_i(x) \tag{12.1a}$$

in the case of optimal plastic design, and by

$$G_0 = G_0(a, x) := -\mu \tag{12.1b}$$

in the second case of limit load analysis.

Here, $x = (x_1, x_2, \ldots, x_r)^T, x := (x_1) = (\mu)$ is the decision vector, hence, the r-vector x of design variables $x_1, \ldots, x_r$, such as sizing variables, in the first case and the load factor $x_1 = \mu$ in the second case. For the decision vector x one has mostly simple feasibility conditions represented by $x \in D$, where $D \subset \mathbb{R}^r$ is a given closed convex set such as $D = \mathbb{R}_+$ in the second case. Moreover, $a = a(\omega)$ is the ν-vector of all random model parameters arising in the underlying mechanical model, such as weight or cost factors $\gamma_{i0} = \gamma_{i0}(\omega)$, yield stresses in compression and tension $\sigma_{yi}^L = \sigma_{yi}^L(\omega), \sigma_{yi}^U = \sigma_{yi}^U(\omega), i = 1, \ldots, B$, load factors contained in the external loading $P = P\Big(a(\omega), x\Big)$, etc. Furthermore, in the general cost function defined by (12.1a), $A_i = A_i(x), i = 1, \ldots, B$, denote the cross-sectional areas of the bars having length $L_i, i = 1, \ldots, B$.

As already mentioned above, the optimization of the function $G_0 = G_0(a, x)$ is done under the safety or survival conditions of plasticity theory which can be described [6, 10] for plane frames as follows:

(I) *Equilibrium condition*
After taking into account the boundary conditions, the equilibrium between the m-vector of external loads $P = P\Big(a(\omega), x\Big)$ and the $3B$-vector of internal loads $F = (F_1^T, F_2^T, \ldots, F_B^T)^T$ can be described by

$$CF = P\Big(a(\omega), x\Big), \tag{12.2}$$

where C is the $m \times 3B$ equilibrium matrix having rank $C = m$.

(II) *Yield condition (feasibility condition)*
If no interactions between normal (axial) forces t_i and bending moments m_i^l, m_i^r, resp., at the left, right end of the oriented i-th bar of the structure are taken into account, then the feasibility condition for the generalized forces of the bar

$$F_i = \begin{pmatrix} t_i \\ m_i^l \\ m_i^r \end{pmatrix}, \quad i = 1, \ldots, B \tag{12.3}$$

reads

$$\widetilde{F}_i^L\Big(a(\omega), x\Big) \le F_i \le \widetilde{F}_i^U\Big(a(\omega, x\Big), \quad i = 1, \ldots, B, \tag{12.4a}$$

where the bounds $\widetilde{F}_i^L, \widetilde{F}_i^U$ containing the plastic capacities with respect to axial forces and moments are given by

$$\widetilde{F}_i^L\Big(a(\omega),x\Big) := \begin{pmatrix} -N_{ipl}^L\Big(a(\omega),x\Big) \\ -M_{ipl}\Big(a(\omega),x\Big) \\ -M_{ipl}\Big(a(\omega),x\Big) \end{pmatrix} = \begin{pmatrix} \sigma_{yi}^L\Big(a(\omega)\Big)A_i(x) \\ -\sigma_{yi}^U\Big(a(\omega)\Big)W_{ipl}(x) \\ -\sigma_{yi}^U\Big(a(\omega)\Big)W_{ipl}(x) \end{pmatrix} \tag{12.4b}$$

$$\widetilde{F}_i^U\Big(a(\omega),x\Big) = \begin{pmatrix} N_{ipl}^U\Big(a(\omega),x\Big) \\ M_{ipl}\Big(a(\omega),x\Big) \\ M_{ipl}\Big(a(\omega),x\Big) \end{pmatrix} = \begin{pmatrix} \sigma_{yi}^U\Big(a(\omega)\Big)A_i(x) \\ \sigma_{yi}^U\Big(a(\omega)\Big)W_{ipl}(x) \\ \sigma_{yi}^U\Big(a(\omega)\Big)W_{ipl}(x) \end{pmatrix}. \tag{12.4c}$$

Here,

$$W_{ipl} = A_i\,\bar{y}_{ic} \tag{12.4d}$$

denotes the plastic section modulus with the arithmetic mean

$$\overline{y}_{ic} = \frac{1}{2}(y_{i1} + y_{i2}) \tag{12.4e}$$

of the centroids y_{i1}, y_{i2} of the two half areas of the cross-sectional areas A_i of the bars $i = 1, \ldots, B$.
Taking into account also interactions between normal forces t_i and moments m_i^l, m_i^r, besides (12.4a) we have additional feasibility conditions of the type

$$-h_l\eta_i^L\Big(a(\omega),x\Big) \le H_l^{(i)}F_i \le h_l\eta_i^U\Big(a(\omega),x\Big), \tag{12.4f}$$

where $\Big(H_l^{(i)}(N_{i0},M_{i0}),h_l\Big)$, $l = 4,\ldots,l_0+3$, are given row vectors depending on the yield domains of the bars, and η_i^L, η_i^U are defined by

$$\eta_i^L\Big(a(\omega),x\Big) = \min\left\{\frac{N_{ipl}^L\Big(a(\omega),x\Big)}{N_{i0}}, \frac{M_{ipl}\Big(a(\omega),x\Big)}{M_{i0}}\right\} \tag{12.4g}$$

$$\eta_i^U\Big(a(\omega),x\Big) = \min\left\{\frac{N_{ipl}^U\Big(a(\omega),x\Big)}{N_{io}}, \frac{M_{ipl}\Big(a(\omega),x\Big)}{M_{i0}}\right\} \tag{12.4h}$$

with certain chosen reference values $N_{i0}, M_{i0}, i = 1,\ldots,B$, for the plastic capacities.

According to (12.4a), (12.4f), the feasibility condition for the vector F of interior loads (member forces and moments) can be represented uniformly by the conditions

$$F_{il}^{L}\Big(a(\omega),x\Big) \le H_{l}^{(i)} F_i \le F_{il}^{U}\Big(a(\omega),x\Big), \quad i=1,\ldots,B,\ l=1,2,\ldots,l_0+3, \tag{12.5a}$$

where the row 3-vectors $H_l^{(i)}$ and the bounds F_{il}^{L}, F_{il}^{U}, $i=1,\ldots,B, l=1,\ldots,l_0+3$, are defined by (12.4a)–(12.4c) and (12.4f)–(12.4h). Let $e_1 =: H_1^T$, $e_2 =: H_2^T$, $e_3 =: H_3^T$ denote the unit vectors of $\mathbb{R}^3$.

Defining the $(l_0+3)\times 3$ matrix $H^{(i)}$ by

$$H^{(i)} := \begin{pmatrix} e_1^T \\ e_2^T \\ e_3^T \\ H_4(N_{i0}, M_{i0}) \\ \vdots \\ H_{l_0+3}(N_{i0}, M_{i0}) \end{pmatrix} \tag{12.5b}$$

and the (l_0+3)-vectors $F_i^L = F_i^L\Big(a(\omega),x\Big)$, $F_i^U = F_i^U\Big(a(\omega),x\Big)$ by

$$F_i^L := \begin{pmatrix} \widetilde{F}_i^L \\ -h_1\eta_i^L \\ \vdots \\ -h_{l_0}\eta_i^L \end{pmatrix}, \quad F_i^U := \begin{pmatrix} \widetilde{F}_i^U \\ h_1\eta_i^U \\ \vdots \\ h_{l_0}\eta_i^U \end{pmatrix}, \tag{12.5c}$$

the feasibility condition can also be represented by

$$F_i^L\Big(a(\omega),x\Big) \le H^{(i)} F_i \le F_i^U\Big(a(\omega),x\Big), \quad i=1,\ldots,B. \tag{12.6}$$

12.2 State and Cost Functions

Defining the quantities

$$F_{il}^{c} = F_{il}^{c}\Big(a(\omega),x\Big) := \frac{F_{il}^{L}+F_{il}^{U}}{2} \tag{12.7a}$$

$$\varrho_{il} = \varrho_{il}\Big(a(\omega),x\Big) := \frac{F_{il}^{U}-F_{il}^{L}}{2}, \tag{12.7b}$$

$i=1,\ldots,B, l=1,\ldots,l_0+3$, the feasibility condition (12.5a) or (12.6) can be described by

$$|z_{il}| \le 1, \quad i=1,\ldots,B,\ l=1,\ldots,l_0+3, \tag{12.8a}$$

with the quotients

$$z_{il} = z_{il}\big(F_i; a(\omega), x\big) = \frac{H_l^{(i)} F_i - F_{il}^c}{\varrho_{il}}, \quad i = 1, \ldots, B,\ l = 1, \ldots, l_0 + 3. \tag{12.8b}$$

The quotient $z_{il}, i = 1, \ldots, B, l = 1, \ldots, l_0 + 3$, denotes the relative deviation of the load component $H_l^{(i)} F_i$ from its "ideal" value F_{il}^c with respect to the radius ϱ_{il} of the feasible interval $[F_{il}^L, F_{il}^U]$. According to (12.8a), (12.8b), the absolute values $|z_{il}|$ of the quotients z_{il} should not exceed the value 1. The absolute value $|z_{il}|$ of the quotient z_{il} denotes the percentage of use of the available plastic capacity by the corresponding load component. Obviously, $|z_{il}| = 1$, $|z_{il}| > 1$, resp., means maximal use, overcharge of the available resources.

Consider now the $(l_0 + 3)$-vectors

$$z_i := (z_{i1}, z_{i2}, \ldots, z_{il_0+3})^T = \left(\frac{H_1^{(i)} F_i - F_{i1}^c}{\varrho_{i1}}, \frac{H_2^{(i)} F_i - F_{i2}^c}{\varrho_{i2}}, \cdots, \frac{H_{l_0+3}^{(i)} F_i - F_{il_0+3}^c}{\varrho_{il_0+3}} \right)^T. \tag{12.8c}$$

With

$$\varrho_i := \begin{pmatrix} \varrho_{i1} \\ \varrho_{i2} \\ \vdots \\ \varrho_{il_0+3} \end{pmatrix}, \quad \varrho_{id} := \begin{pmatrix} \varrho_{i1} & 0 & \ldots & 0 \\ 0 & \varrho_{i2} & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \varrho_{il_0+3} \end{pmatrix}, \quad F_i^c := \begin{pmatrix} F_{i1}^c \\ F_{i2}^c \\ \vdots \\ F_{il_0+3}^c \end{pmatrix} \tag{12.8d}$$

we get

$$z_i = \varrho_{id}^{-1}(H^{(i)} F_i - F_i^c). \tag{12.8e}$$

Using (12.4b)–(12.4d), we find

$$F_i^c = \left(A_i \frac{\sigma_{yi}^L + \sigma_{yi}^U}{2}, 0, 0, h_4 \frac{\eta_i^U - \eta_i^L}{2}, \ldots, h_{l_0+3} \frac{\eta_i^U - \eta_i^L}{2} \right)^T \tag{12.8f}$$

$$\varrho_i = \left(A_i \frac{\sigma_{yi}^U - \sigma_{yi}^L}{2}, A_i \sigma_{yi}^U \bar{y}_{ic}, A_i \sigma_{yi}^U \bar{y}_{ic}, h_4 \frac{\eta_i^U + \eta_i^L}{2}, \ldots, h_{l_0+3} \frac{\eta_i^U + \eta_i^L}{2} \right)^T. \tag{12.8g}$$

The vector z_i can be represented then, cf. (12.3), by

$$z_i = \left(\frac{t_i - A_i \dfrac{\sigma_{yi}^L + \sigma_{yi}^U}{2}}{A_i \dfrac{\sigma_{yi}^U - \sigma_{yi}^L}{2}}, \frac{m_i^l}{A_i \sigma_{yi}^U \bar{y}_{ic}}, \frac{m_i^r}{A_i \sigma_{yi}^U \bar{y}_{ic}}, \frac{H_4^{(i)} F_i - h_4 \frac{\eta_i^U - \eta_i^L}{2}}{h_4 \frac{\eta_i^U + \eta_i^L}{2}}, \ldots, \right.$$
$$\left. \frac{H_{l_0+3}^{(i)} F_i - h_{l_0+3} \frac{\eta_i^U - \eta_i^L}{2}}{h_{l_0+3} \frac{\eta_i^U + \eta_i^L}{2}} \right)^T . \tag{12.9a}$$

In case of symmetry $\sigma_{yi}^L = -\sigma_{yi}^U$ we get

$$z_i = \left(\frac{t_i}{A_i \sigma_{y_i}^U}, \frac{m_i^l}{A_i \sigma_{y_i}^U \bar{y}_{ic}}, \frac{m_i^r}{A_i \sigma_{y_i}^U \bar{y}_{ic}}, \frac{H_4^{(i)} F_i}{h_4 \eta_i^U}, \ldots, \frac{H_{l_0+3}^{(i)} F_i}{h_{l_0+3} \eta_i^U} \right)^T . \tag{12.9b}$$

According to the methods introduced in [6–8], the fulfillment of the survival condition for elastoplastic frame structures, hence, the equilibrium condition (12.2) and the feasibility condition (12.6) or (12.8a), (12.8b), can be described by means of the *state function* $s^* = s^*\Big(a(\omega), x\Big)$ defined, in the present case, by

$$s^* = s^*\Big(a(\omega), x\Big) := \min \Big\{ s : \Big| z_{il}\Big(F_i; a(\omega), x\Big) \Big| - 1 \le s, i = 1, \ldots, B,$$
$$l = 1, 2, \ldots, l_0 + 3, CF = P\Big(a(\omega), x\Big) \Big\} . \tag{12.10}$$

Hence, the state function s^* is the minimum value function of the linear program (LP)

$$\min s \tag{12.11a}$$

s.t.

$$\Big| z_{il}\Big(F_i; a(\omega), x\Big) \Big| - 1 \le s, \quad i = 1, \ldots, B, \quad l = 1, \ldots, l_0 + 3 \tag{12.11b}$$

$$CF = P\Big(a(\omega), x\Big). \tag{12.11c}$$

Since the objective function s is bounded from below and a feasible solution (s, F) always exists, LP (12.11a)–(12.11c) has an optimal solution $(s^*, F^*) = \Big(s^*\Big(a(\omega), x\Big), F^*\Big(a(\omega), x\Big)\Big)$ for each configuration $\Big(a(\omega), x\Big)$ of the structure.

Consequently, for the survival of the structure we have the following criterion, cf. [7].

Theorem 12.1 *The elastoplastic frame structure having configuration (a, x) carries the exterior load $P = P(a, x)$ safely if and only if*

$$s^*(a,x) \le 0. \tag{12.12}$$

Obviously, the constraint (12.11b) in the LP (12.11a)–(12.11c) can also be represented by

$$\left\| z\Big(F; a(\omega), x\Big) \right\|_\infty - 1 \le s, \tag{12.13a}$$

where $z = z\Big(F; a(\omega), x\Big)$ denotes the $B(l_0+3)$-vector

$$z\Big(F; a(\omega), x\Big) := \left(z_1\Big(F; a(\omega), x\Big)^T, \ldots, z_B\Big(F; a(\omega), x\Big)^T \right)^T, \tag{12.13b}$$

and $\|z\|_\infty$ is the maximum norm

$$\|z\|_\infty := \max_{\substack{1\le i\le B \\ 1\le l\le l_0+3}} |z_{il}|. \tag{12.13c}$$

If we put

$$\hat{s} = 1 + s \quad \text{or} \quad s = \hat{s} - 1, \tag{12.14}$$

from (12.10) we obtain

$$s^*(a,x) = \hat{s}^*(a,x) - 1, \tag{12.15a}$$

where the transformed state function $\hat{s}^* = \hat{s}^*(a,x)$ reads

$$\hat{s}^*(a,x) := \min \left\{ \left\| z(F; a, x) \right\|_\infty : CF = P(a,x) \right\}. \tag{12.15b}$$

Remark 12.1 According to (12.15a), (12.15b) and (12.12), the safety or survival condition of the plane frame with plastic material can be represented also by

$$\hat{s}^*(a,x) \le 1.$$

The state function $\hat{s}^* = \hat{s}^*(a,x)$ describes the maximum percentage of use of the available plastic capacity within the plane frame for the best internal load distribution with respect to the configuration (a,x).

Obviously, $\hat{s}^* = \hat{s}^*(a,x)$ is the minimum value function of the LP

$$\min_{CF=P(a,x)} \left\| z(F; a, x) \right\|_\infty. \tag{12.16}$$

The following inequalities for norms or power/Hölder means $\|z\|$ in $\mathbb{R}^{B(l_0+3)}$ are well known [1, 4]:

$$\frac{1}{B(l_0+3)}\|z\|_\infty \le \frac{1}{B(l_0+3)}\|z\|_1$$
$$\le \frac{1}{\sqrt{B(l_0+3)}}\|z\|_2 \le \|z\|_\infty \le \|z\|_2, \tag{12.17a}$$

where

$$\|z\|_1 := \sum_{i=1}^{B}\sum_{l=1}^{l_0+3}|z_{il}|,\ \|z\|_2 := \sqrt{\sum_{i=1}^{B}\sum_{l=1}^{l_0+3} z_{il}^2}. \tag{12.17b}$$

Using (12.17a), we find

$$\frac{1}{B(l_0+3)}\hat{s}^*(a,x) \le \frac{1}{\sqrt{B(l_0+3)}}\hat{s}_2^*(a,x) \le \hat{s}^*(a,x) \le \hat{s}_2^*(a,x), \tag{12.18a}$$

where $\hat{s}_2^* = \hat{s}_2^*(a,x)$ is the modified state function defined by

$$\hat{s}_2^*(a,x) := \min\left\{\left\|z(F;a,x)\right\|_2 : CF = P(a,x)\right\}. \tag{12.18b}$$

Obviously, we have

$$\hat{s}_2^*(a,x) = \sqrt{G_1^*(a,x)}, \tag{12.18c}$$

where $G_1^*(a,x)$ is the minimum value function of the **quadratic program**

$$\min_{CF=P\left(a(\omega),x\right)} \sum_{i=1}^{B}\sum_{l=1}^{l_0+3} z_{il}(F_i;a,x)^2. \tag{12.19}$$

12.2.1 Cost Functions

The inequalities in (12.18a) show that for structural analysis and optimal design purposes we may work also with the state function $\hat{s}_2^* = \hat{s}_2^*(a,x)$ which can be defined easily by means of the quadratic program (12.19).

According to the definition (12.8b) and the corresponding technical interpretation of the quotients z_{il}, the transformed state function $\hat{s}^* = \hat{s}^*(a,x)$ represents—for the best internal load distribution—the maximum percentage of use of the plastic capacities *relative* to the available plastic capacities in the members (bars) of the plane frame with configuration (a,x). While the definition (12.15b) of $\hat{s}^*$ is based on the absolute value function

$$c_1(z_{il}) = |z_{il}|, \tag{12.20a}$$

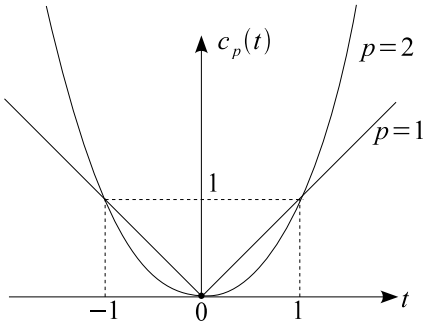


Fig. 12.1 Cost functions c_p

in definition (12.18b) of $\hat{s}_2^*$ occur quadratic functions

$$c_2(z_{il}) = z_{il}^2, \quad i = 1, \ldots, B, \quad l = 1, \ldots, l_0 + 3. \tag{12.20b}$$

Obviously,

$$c_p(z_{il}) = |z_{il}|^p \text{ with } p = 1, 2 \text{ (or also } p \geq 1)$$

are possible convex functions, cf. Fig. 12.1, measuring the **costs** resulting from the position z_{il} of a load component $\widetilde{H}_l^{(i)} F_i$ relative to the corresponding safety interval (plastic capacity) $[\widetilde{F}_{il}^L, \widetilde{F}_{il}^U]$.

If different weights are used in the objective function (12.19), then for the bars we obtain, cf. (12.8c), the cost functions

$$q_i(z_i) = \| W_{i0} z_i \|^2, \tag{12.20c}$$

with $(l_0 + 3) \times (l_0 + 3)$ weight matrices W_{i0}, $i = 1, \ldots, B$.

The total weighted quadratic costs resulting from a load distribution F acting on the plastic plane frame having configuration (a, x) are given, cf. (12.18c), (12.19), (12.20c), by

$$G_1 := \sum_{i=1}^{B} \| W_{i0} z_i \|^2 = \sum_{i=1}^{B} z_i^T W_{i0}^T W_{i0} z_i. \tag{12.21a}$$

Defining

$$W_0 := \begin{pmatrix} W_{10} & 0 & \ldots & 0 \\ 0 & W_{20} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & W_{B0} \end{pmatrix}, \quad z := \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_B \end{pmatrix}, \tag{12.21b}$$

we also have

$$G_1 = G_1(a, x; F) = z^T W_0^T W_0 z$$
$$= \|W_0 z\|_2^2 = \|z\|_{2,W_0}^2, \tag{12.21c}$$

where $\|\cdot\|_{2,W_0}$ denotes the *weighted Euclidean norm*

$$\|z\|_{2,W_0} := \|W_0 z\|_2. \tag{12.21d}$$

Using the weighted quadratic cost function (12.20c), the state function $\hat{s}_2^* = \hat{s}_2^*(a, x)$ is replaced by

$$\hat{s}_{2,W_0}^*(a, x) := \min\left\{\left\|z(F; a, x)\right\|_{2,W_0} : CF = P(a, x)\right\}$$
$$= \min\left\{\sqrt{G_1(a, x; F)} : CF = P(a, x)\right\}. \tag{12.21e}$$

Since

$$\|z\|_{2,W_0} = \|W_0 z\|_2 \le \|W_0\| \cdot \|z\|$$

with the norm $\|W_0\|$ of the matrix W_0, we find

$$\hat{s}_{2,W_0}^*(a, x) \le \|W_0\| \hat{s}_2^*(a, x). \tag{12.21f}$$

On the other hand, in case

$$\|W_0 z\|_2 \ge \underline{W}_0 \|z\|_2$$

with a positive constant $\underline{W}_0 > 0$, we have

$$\hat{s}_{2,W_0}^*(a, x) \ge \underline{W}_0 \hat{s}_2^*(a, x) \text{ or } \hat{s}_2^*(a, x) \le \frac{1}{\underline{W}_0} \hat{s}_{2,W_0}^*(a, x). \tag{12.21g}$$

12.3 Minimum Expected Quadratic Costs

Putting

$$H := \begin{pmatrix} H^{(1)} & & & \\ & H^{(2)} & & \\ & & \ddots & \\ & & & H^{(B)} \end{pmatrix}, \quad F^c := \begin{pmatrix} F_1^c \\ F_2^c \\ \vdots \\ F_B^c \end{pmatrix}, \quad \varrho := \begin{pmatrix} \varrho_1 \\ \varrho_2 \\ \vdots \\ \varrho_B \end{pmatrix}, \tag{12.21h}$$

with (12.8e) we find

$$G_1 = G_1\Big(a(\omega), x; F\Big) = (HF - F^c)^T \varrho_d^{-1} W_0^T W_0 \varrho_d^{-1} (HF - F^c). \tag{12.22a}$$

If

$$F^c = HF^c \text{ with } F^c := \left(\frac{F_i^L + F_i^U}{2}\right)_{i=1,\ldots,B} \tag{12.22b}$$

as in the case of no interaction between normal forces and moments, see (12.4a)–(12.4c), and in the case of symmetric yield stresses

$$\sigma_{yi}^L = -\sigma_{yi}^U, \quad i = 1, \ldots, B, \tag{12.22c}$$

we also have

$$G_1\Big(a(\omega), x; F\Big) = (F - F^c)^T H^T \varrho_d^{-1} W_0^T W_0 \varrho_d^{-1} H(F - F^c). \tag{12.22d}$$

Moreover, if (12.22c) holds, then $F^c = 0$ and therefore

$$G_1\Big(a(\omega), x; F\Big) = F^T H^T \varrho_d^{-1} W_0^T W_0 \varrho_d^{-1} HF. \tag{12.22e}$$

For simplification, we assume first in this section that the total cost representation (12.22d) or (12.22e) holds.

According to the equilibrium condition (12.2), the total vector F of generalized forces of the members fulfills

$$CF = P(a(\omega), x).$$

Let $x \in D$ denote a given vector of decision variables, and let be $a = a(\omega)$ a realization of vector $a(\cdot)$ of model parameters. Based on (12.22d) or (12.22e), a cost minimum or "best" internal distribution of the generalized forces

$$F^* = F^*(a(\omega), x)$$

of the structure can be obtained by solving the following optimization problem with quadratic objective function and linear constraints

$$\min_{CF = P(a(\omega), x)} G_1(a(\omega), x; F). \tag{12.23}$$

Solving the related stochastic optimization problem [8]

$$\min_{CF = P(a(\omega), x) \text{a.s.}} EG_1(a(\omega), x; F), \tag{12.24}$$

for the random configuration $\big(a(\omega), x\big)$ we get the minimum expected (total) quadratic costs

$$\overline{G}_1^* = \overline{G}_1^*(x), \ x \in D, \tag{12.25a}$$

where $\overline{G}_1^*(x)$ may be obtained by interchanging expectation and minimization

$$\overline{G}_1^*(x) = \overline{G_1^*}(x) := E \min\{G_1(a(\omega), x; F) : CF = P(a(\omega), x)\}. \tag{12.25b}$$

The internal minimization problem (12.23)

$$\min\ G_1(a(\omega), x; F) \quad \text{s.t. } CF = P(a(\omega), x),$$

hence,

$$\min_{CF=P(a(\omega),x)} (F - F^c)^T H^T \varrho_d^{-1} W_0^T W_0 \varrho_d^{-1} H (F - F^c), \tag{12.26}$$

with quadratic objective function and linear constraints with respect to F can be solved by means of Lagrange techniques. We put

$$W = W(a, x) := H^T \varrho_d^{-1} W_0^T W_0 \varrho_d^{-1} H \tag{12.27}$$

and define the Lagrangian of (12.26):

$$L = L(F, \lambda) := (F - F^c)^T W (F - F^c) + \lambda^T (CF - P(a(\omega), x)). \tag{12.28a}$$

Based on the corresponding piecewise linearized yield domain, W describes the plastic capacity of the plane frame with respect to axial forces and bending moments.

The necessary and sufficient optimality conditions for a minimum point (F^*, λ^*) read

$$0 = \nabla_F L = 2W(F - F^c) + C^T \lambda, \tag{12.28b}$$

$$0 = \nabla_\lambda L = CF - P. \tag{12.28c}$$

Supposing that W is regular, we get

$$F = F^c - \frac{1}{2} W^{-1} C^T \lambda \tag{12.28d}$$

and

$$P = CF = CF^c - \frac{1}{2} C W^{-1} C^T \lambda, \tag{12.28e}$$

hence,

$$F^* = F^c - \frac{1}{2} W^{-1} C^T \lambda^* = F^c - W^{-1} C^T (C W^{-1} C^T)^{-1} (CF^c - P). \tag{12.28f}$$

Inserting (12.28f) into the objective function $G_1(a(\omega), x; F)$, according to (12.22a) and (12.27) we find

$$
\begin{aligned}
G_1^* &= G_1^*(a(\omega), x) \\
&= (F^* - F^c)^T W (F^* - F^c) \\
&= \left((CF^c - P)^T (CW^{-1}C^T)^{-1} CW^{-1}\right) W \left(W^{-1}C^T (CW^{-1}C^T)^{-1}(CF^c - P)\right) \\
&= (CF^c - P)^T (CW^{-1}C^T)^{-1}(CF^c - P) \\
&= tr(CW^{-1}C^T)^{-1}(CF^c - P)(CF^c - P)^T, \qquad (12.28g)
\end{aligned}
$$

where "tr" denotes the trace of a matrix. The minimal expected value $\overline{G_1^*}$ is then given by

$$
\begin{aligned}
\overline{G_1^*}(x) &= E G_1^*\big(a(\omega), x\big) \\
&= E\big(CF^c(a(\omega), x) - P\big(a(\omega), x\big)\big)^T \Big(CW\big(a(\omega), x\big)^{-1} C^T\Big)^{-1} \\
&\quad \times \Big(CF^c\big(a(\omega), x\big) - P\big(a(\omega), x\big)\Big) \\
&= E\, tr\Big(CW\big(a(\omega), x\big)^{-1} C^T\Big)^{-1} \Big(CF^c\big(a(\omega), x\big) - P\big(a(\omega), x\big)\Big) \\
&\quad \times \Big(CF^c\big(a(\omega), x\big) - P\big(a(\omega), x\big)\Big)^T. \qquad (12.29a)
\end{aligned}
$$

If $\sigma_{yi}^L = -\sigma_{yi}^U,\ i = 1, \ldots, B$, then $F^c = 0$ and

$$
\begin{aligned}
\overline{G_1^*}(x) &= E P(a(\omega), x))^T (CW(a(\omega), x)^{-1} C^T)^{-1} P(a(\omega), x) \\
&= tr E\Big(CW\big(a(\omega), x\big)^{-1} C^T\Big)^{-1} P\big(a(\omega), x\big) P\big(a(\omega), x\big)^T. \qquad (12.29b)
\end{aligned}
$$

Since the vector $P = P(a(\omega), x)$ of external generalized forces and the vector of yield stresses $\sigma^U = \sigma^U(a(\omega), x)$ are stochastically independent, then in case $\sigma_{yi}^L = -\sigma_{yi}^U,\ i = 1, \ldots, B$, we have

$$
\begin{aligned}
\overline{G_1^*}(x) &= E P(a(\omega), x))^T \overline{U}(x) P(a(\omega), x) \\
&= tr \overline{U}(x) E P(a(\omega), x) P(a(\omega), x)^T, \qquad (12.29c)
\end{aligned}
$$

where

$$
\overline{U}(x) := E K(a(\omega), x)^{-1} \qquad (12.29d)
$$

with the matrices

$$
K(a, x) := C K_0(a, x) C^T \qquad (12.30a)
$$

and

$$
K_0(a, x) := W(a, x)^{-1} = (H^T \varrho_d^{-1} W_0^T W_0 \varrho_d^{-1} H)^{-1}. \qquad (12.30b)
$$

We compare now, especially in case $F^c = 0$, formula (12.28g) for the costs $G_1^* = G_1^*(a, x)$ with formula

$$\Gamma := u^T P$$

for the *compliance* of an elastic structure, where

$$u := K_{el}^{-1} P$$

is the vector of displacements, and K_{el} denotes the stiffness matrix in case of an elastic structure. Obviously, the cost function $G_1^* = G_1^*(a, x)$ may be interpreted as a *generalized compliance function*, and the $m \times m$ matrix $K = K(a, x)$ can be interpreted as the "*generalized stiffness matrix*" of the underlying plastic mechanical structure. If we suppose that

$$W_{i0} := \left(w_{il}^0 \delta_{l\lambda}\right)_{l,\lambda=1,\ldots,l_0+3}, \quad i = 1, \ldots, B \tag{12.30c}$$

are diagonal weight matrices, then, cf. (12.8g),

$$\varrho_d^{-1} W_0^T W_0 \varrho_d^{-1} = \operatorname{diag}\left(\left(\frac{w_{il}^0}{\varrho_{il}}\right)^2\right). \tag{12.30d}$$

If condition (12.22b) and therefore representation (12.22d) or (12.22e) does not hold, then the minimum total costs $G_1^* = G_1^*\Big(a(\omega), x\Big)$ are determined by the more general quadratic program, cf. (12.22a), (12.23), (12.26),

$$\min_{CF=P\left(a(\omega),x\right)} (HF - F^c)^T W_\varrho\Big(a(\omega), x\Big)(HF - F^c), \tag{12.31a}$$

where

$$W_\varrho(a, x) := \varrho_d^{-1}(a, x) W_0^T W_0 \varrho_d^{-1}(a, x). \tag{12.31b}$$

Though also in this case problem (12.31a) can be solved explicitly, the resulting total cost function has a difficult form. In order to apply the previous technique, the vector F^c is approximated—in the least squares sense—by vectors HF with $F \in \mathbb{R}^{3B}$. Hence, we write

$$F^c \approx HF^{c*}, \tag{12.32a}$$

where the $(3B)$-vector F^{c*} is the optimal solution of the optimization problem

$$\min_F \|HF - F^c\|^2. \tag{12.32b}$$

We obtain

$$F^{c*} = F^{c*}(F^c) := (H^T H)^{-1} H^T F^c. \tag{12.32c}$$

The error $e(F^{c*})$ of this approximation reads

$$\begin{aligned} e(F^c) &:= \|HF^{c*} - F^c\| \\ &= \left\| \left(H(H^T H)^{-1} H^T - I \right) F^c \right\|. \end{aligned} \tag{12.32d}$$

With the vector $F^{c*} = F^{c*}(F^c)$ the total costs $G_1^a = G_1^a\Big(a(\omega), x; F\Big)$ can be approximated now, see (12.22a), by

$$\begin{aligned} G_1^a\Big(a(\omega), x; F\Big) &:= (HF - HF^{c*})^T W_\varrho\Big(a(\omega), x\Big)(HF - HF^{c*}) \\ &= (F - F^{c*})^T H^T W_\varrho\Big(a(\omega), x\Big) H(F - F^{c*}) \\ &= (F - F^{c*})^T W\Big(a(\omega), x\Big)(F - F^{c*}), \end{aligned} \tag{12.33a}$$

where, cf. (12.27),

$$W = W(a, x) := H^T W_\varrho(a, x) H. \tag{12.33b}$$

Obviously, the approximate cost function $G_1^a = G_1^a(a, x; F)$ has the same form as the cost function $G_1 = G_1(a, x; F)$ under the assumption (12.22b), see (12.22d). Hence, the minimum cost function $G_1^{a*} = G_1^{a*}(a, x)$ can be determined by solving, cf. (12.23),

$$\min_{CF = P(a,x)} G_1^a(a, x; F). \tag{12.34a}$$

We get, see (12.28g),

$$G_1^{a*}(a, x) := tr\Big(CW(a, x)^{-1} C^T\Big)^{-1} (CF^{c*} - P)(CF^{c*} - P)^T, \tag{12.34b}$$

where $F^{c*} = F^{c*}(a, x)$ is given here by (12.32c). Taking expectations in (12.34b), we obtain the approximate minimum expected total cost function

$$\overline{G_1^{a*}} = \overline{G_1^{a*}}(x) = E G_1^{a*}\Big(a(\omega), x\Big). \tag{12.34c}$$

12.4 Deterministic Substitute Problems

In order to determine robust optimal designs x^*, appropriate deterministic substitute problems, cf. [8], must be formulated.

12.4.1 Weight (Volume)-Minimization Subject to Expected Cost Constraints

With the expected primary cost function, see (12.1a), (12.1b),

$$\overline{G_0}(x) = EG_0(a(\omega), x)$$

and the expected cost function $\overline{G_1^*} = \overline{G_1^*}(x)$ representing the expected total weighted quadratic costs resulting from a violation of the feasibility condition (12.4a), (12.4f), we get [3, 9] the optimization problem

$$\min \ \overline{G_0}(x) \tag{12.35a}$$

$$\text{s.t. } \overline{G_1^*}(x) \le \Gamma_1 \tag{12.35b}$$

$$x \in D, \tag{12.35c}$$

where Γ_1 is a certain upper cost bound. In case (12.1a) we have

$$\overline{G_0}(x) := \sum_{i=1}^{B} \overline{\gamma_{i0}} L_i A_i(x) \tag{12.35d}$$

with $\overline{\gamma_{i0}} := E\gamma_{i0}(\omega)$, and $\overline{G_1^*} = \overline{G_1^*}(x)$ is defined by (12.29a) or (12.29b).

Due to (12.20c) and (12.21a)–(12.21c), the upper cost bound Γ_1 can be defined by

$$\Gamma_1 := g_1 G_1^{\max}, \tag{12.35e}$$

where $g_1 > 0$ is a certain reliability factor, and $G_1^{\max}$ denotes the maximum of the total cost function $G_1 = G_1(z)$ on the total admissible z-domain $[-1, 1]^{(l_0+3)B}$. Hence,

$$\begin{aligned} G_1^{\max} &:= \max_{z \in [-1,1]^{(l_0+3)B}} \sum_{i=1}^{B} \| W_{i0} z_i \|^2 \\ &= \sum_{i=1}^{B} \max_{z_i \in [-1,1]^{(l_0+3)}} \| W_{i0} z_i \|^2 \\ &= \sum_{i=1}^{B} \max_{1 \le j \le 2^{l_0+3}} \| W_{i0} e^{(j)} \|^2, \end{aligned} \tag{12.35f}$$

where $e^{(j)}$, $j = 1, \ldots, 2^{l_0+3}$, denote the extreme points of the hypercube $[-1, 1]^{l_0+3}$.

As shown in the following, for $W_0 = I$ (identity matrix) the expected cost constraint (12.35b) can also be interpreted as a reliability constraint.

According to Theorem 12.1, (12.12) and (12.15a), (12.15b), for the probability of survival $p_s = p_s(x)$ of the elastoplastic structure represented by the design vector x we have

$$p_s(x) := P\Big(s^*\big(a(\omega), x\big) \le 0\Big)$$
$$= P\Big(\hat{s}^*\big(a(\omega), x\big) - 1 \le 0\Big) = P\Big(\hat{s}^*\big(a(\omega), x\big) \le 1\Big). \tag{12.36}$$

Knowing from (12.18a), (12.18b) that, in case $W_0 = I$,

$$\frac{1}{\sqrt{B(l_0+3)}}\hat{s}_2^*(a,x) \le \hat{s}^*(a,x) \le \hat{s}_2^*(a,x),$$

we obtain the probability inequalities

$$P\Big(\hat{s}_2^*\big(a(\omega), x\big) \le 1\Big) \le p_s(x) \le P\Big(\hat{s}_2^*(a,x) \le \sqrt{B(l_0+3)}\Big). \tag{12.37a}$$

Due to the first definition of $G_1^* = G_1^*(a,x)$ by (12.18c) and (12.19), related to the case $W_0 = I$, we also have

$$P\Big(G_1^*\big(a(\omega), x\big) \le 1\Big) \le p_s(x) \le P\Big(G_1^*\big(a(\omega), x\big) \le B(l_0+3)\Big). \tag{12.37b}$$

Using now a nonnegative, nondecreasing, measurable function h on $\mathbb{R}_+$, for any $g_1 > 0$ we find [8]

$$P\Big(G_1^*\big(a(\omega), x\big) \le g_1\Big) \ge 1 - \frac{Eh\Big(G_1^*\big(a(\omega), x\big)\Big)}{h(g_1)}. \tag{12.38a}$$

In the case $h(t) = t$ we get the inequality

$$P\Big(G_1^*\big(a(\omega), x\big) \le g_1\Big) \ge 1 - \frac{\overline{G_1^*}(x)}{g_1}, \tag{12.38b}$$

where the expectation $\overline{G_1^*}(x) = EG_1^*\big(a(\omega), x\big)$ is given by (12.29a) or (12.29b). The probabilistic constraint

$$P\Big(G_1^*\big(a(\omega), x\big) \le g_1\Big) \ge \alpha_{\min} \tag{12.39a}$$

for the quadratic mean rate $\hat{s}_2^* = \sqrt{G_1^*(a,x)}$ of minimum possible use of plastic capacity within the plane frame with configuration (a,x) implies $p_s(x) \ge \alpha_{\min}$ for $g_1 = 1$, cf. (12.37b). Hence, due to (12.38b), constraint (12.39a) and therefore $p_s(x) \ge \alpha_{\min}$ can be guaranteed then by the condition

$$\overline{G_1^*}(x) \le g_1(1-\alpha), \tag{12.39b}$$

see (12.35b).

12.4.2 Minimum Expected Total Costs

For a vector $x \in D$ of decision variables and a vector F of internal generalized forces fulfilling the equilibrium condition (12.2), from (12.1a), (12.1b) and (12.22a), (12.22b) we have the total costs

$$G(a(\omega), x; F) := G_0(a(\omega), x) + G_1(a(\omega), x; F). \tag{12.40a}$$

Here, the weight or scale matrices W_{i0} and the weight or cost factors γ_{i0}, $i = 1, \ldots, B$, must be selected such that the dimensions of G_0 and G_1 coincide. For example, if $W_{i0} = I, i = 1, \ldots, B$, and $\sqrt{G_1(a, x)}$ is then the quadratic mean rate of use of plastic capacity for a given distribution of generalized forces F, then we may replace γ_{i0} by the relative weight/cost coefficients

$$\gamma_{i0}^{rel} := \frac{\gamma_{i0}}{G_0^{ref}}, \quad i = 1, \ldots, B,$$

with a certain weight or cost reference value G_0^{ref}.

Minimizing now the expected total costs

$$\begin{aligned} \overline{G} = \overline{G}(x) &= EG(a(\omega), x; F(\omega)) \\ &= E(G_0(a(\omega), x) + G_1(a(\omega), x; F(\omega))) \\ &= EG_0(a(\omega), x) + EG_1(a(\omega), x; F(\omega)) \\ &= \overline{G}_0(x) + EG_1(a(\omega), x; F(\omega)) \end{aligned} \tag{12.40b}$$

subject to the equilibrium conditions (12.2) and the remaining condition for the decision variables

$$x \in D, \tag{12.40c}$$

we obtain the stochastic optimization problem

$$\min_{\substack{CF(\omega)=P\left(a(\omega),x\right) \text{ a.s.} \\ x \in D}} E(G_0(a(\omega), x) + G_1(a(\omega), x; F(\omega))). \tag{12.41}$$

Obviously, (12.41) has the following *two-stage structure*:

Step 1: Select $x \in D$ without knowledge of the actual realization $a = a(\omega)$ of the model parameters, but knowing the probability distribution or certain moments of $a(\cdot)$;

Step 2: Determine the best internal distribution of generalized forces $F = F^*(\omega)$ after realization of $a = a(\omega)$.

Therefore, problem (12.41) is equivalent to

$$\min_{x\in D} E\left(G_0(a(\omega),x)+\min_{CF=P(a(\omega),x)} G_1(a(\omega),x;F)\right). \tag{12.42}$$

According to the definitions (12.35d) of $\overline{G_0}$ and (12.25b) of $\overline{G_1^*}$, problem (12.42) can be represented also by

$$\min_{x\in D}\left(\overline{G_0}(x)+\overline{G_1^*}(x)\right). \tag{12.43}$$

12.5 Stochastic Nonlinear Programming

We first suppose that the structure consists of a uniform material with a symmetric random yield stress in compression and tension. Hence, we assume next to

$$\sigma_{yi}^U=-\sigma_{yi}^L=\sigma_y^U=\sigma_y^U(\omega),\quad i=1,\dots,B, \tag{12.44}$$

with a random yield stress $\sigma_y^U(\omega)$. Due to (12.8e) we have

$$\begin{aligned}\varrho_i\big(a(\omega),x\big)&=A_i(\sigma_{yi}^U,\sigma_{yi}^U\overline{y}_{ic},\sigma_{yi}^U\overline{y}_{ic},\sigma_{yi}^U h_4\eta_i,\dots,\sigma_{yi}^U h_{l_0+3}\eta_i)^T\\&=\sigma_y^U A_i(1,\overline{y}_{ic},\overline{y}_{ic},h_4\eta_i,\dots,h_{l_0+3}\eta_i)^T:=\sigma_y^U\mathring{\varrho}_i(x)\end{aligned} \tag{12.45a}$$

and therefore, see (12.8d),

$$\varrho\big(a(\omega),x\big)=\sigma_y^U(\omega)\mathring{\varrho}(x) \tag{12.45b}$$

with $\mathring{\varrho}_i(x):=A_i(1,\overline{y}_{ic},\overline{y}_{ic},h_4\eta_i,\dots,h_{l_0+3}\eta_i)^T$, $\eta_i:=\min\left\{\dfrac{1}{N_{i0}},\dfrac{\overline{y}_{ic}}{M_{i0}}\right\}$ and

$$\mathring{\varrho}(x)=\begin{pmatrix}\mathring{\varrho}_1(x)\\ \vdots\\ \mathring{\varrho}_B(x)\end{pmatrix}. \tag{12.45c}$$

According to (12.30a), (12.30b), for *fixed weight matrices* W_{i0}, $i=1,\dots,B$, we obtain

$$K\big(a(\omega),x\big)=CK_0\big(a(\omega),x\big)C^T \tag{12.46a}$$

with

$$K_0\big(a(\omega),x\big) = (H^T\varrho_d^{-1}W_0^T W_0\varrho_d^{-1}H)^{-1}$$
$$= \sigma_y^U(\omega)^2\mathring{K}_0(x), \tag{12.46b}$$

where

$$\mathring{K}_0(x) := \big(H^T\mathring{\varrho}(x)_d^{-1}W_0^T W_0\mathring{\varrho}(x)_d^{-1}H\big)^{-1}. \tag{12.46c}$$

Now, (12.30a), (12.30b), and (12.46a)–(12.46c) yield

$$K\big(a(\omega),x\big) = \sigma_y^U(\omega)^2\, C\mathring{K}_0(x)C^T = \sigma_y^U(\omega)^2\mathring{K}(x) \tag{12.47a}$$

with the deterministic matrix

$$\mathring{K}(x) := C\mathring{K}_0(x)C^T. \tag{12.47b}$$

Moreover, we get

$$U\big(a(\omega),x\big) := K\big(a(\omega),x\big)^{-1} = \big(\sigma_y^U(\omega)^2\mathring{K}(x)\big)^{-1}$$
$$= \frac{1}{\sigma_y^U(\omega)^2}\mathring{K}(x)^{-1}. \tag{12.47c}$$

Hence, see (12.29d),

$$\overline{U}(x) = EU\big(a(\omega),x\big) = \left(E\frac{1}{\sigma_y^U(\omega)^2}\right)\mathring{K}(x)^{-1}. \tag{12.47d}$$

In case of a *random* weight matrix $W_0 = W_0\Big(a(\omega)\Big)$, for $\overline{U}(x)$ we also obtain a representation of the type (12.47d), provided that i) the random variables $W_0\Big(a(\omega)\Big)$ and $\sigma_y^U(\omega)$ are stochastically independent and ii) $\mathring{K}(x)$ is defined by

$$\mathring{K}(x) := \left(E\left(C\mathring{K}_0\Big(W\big(a(\omega)\big),x\Big)C^T\right)^{-1}\right)^{-1}.$$

From (12.29c) we obtain

$$\overline{G_1^*}(x) = EG_1^*\big(a(\omega),x\big)$$
$$= tr\overline{U}(x)EP\big(a(\omega),x\big)P\big(a(\omega),x\big)^T$$
$$= \left(E\frac{1}{\sigma_y^U(\omega)^2}\right)tr\mathring{K}(x)^{-1}EP\big(a(\omega),x\big)P\big(a(\omega),x\big)^T. \tag{12.48}$$

Representing the $m\times m$ matrix

$$\begin{aligned} B(x) &:= E\,P\big(a(\omega),x\big)P\big(a(\omega)x\big)^T \\ &= \overline{P}(x)\overline{P}(x)^T + \operatorname{cov}\Big(P\big(a(\cdot),x\big)\Big) \\ &= \big(b_1(x), b_2(x), \ldots, b_m(x)\big) \end{aligned} \tag{12.49a}$$

by its columns $b_j(x)$, $j = 1, \ldots, m$, where we still set

$$\overline{P}(x) := E\,P\big(a(\omega),x\big) \tag{12.49b}$$

$$\operatorname{cov}\Big(P\big(a(\cdot),x\big)\Big) := E\Big(P\big(a(\omega),x\big) - \overline{P}(x)\Big)\Big(P\big(a(\omega),x\big) - \overline{P}(x)\Big)^T, \tag{12.49c}$$

we find

$$\begin{aligned} Z(x) = (z_1, z_2, \ldots, z_m) &:= E\left(\frac{1}{\sigma_y^U(\omega)^2}\right)\mathring{K}(x)^{-1}B(x) \\ &= E\left(\frac{1}{\sigma_y^U(\omega)^2}\right)\big(\mathring{K}(x)^{-1}b_1(x), \mathring{K}(x)^{-1}b_2(x), \ldots, \mathring{K}(x)^{-1}b_m(x)\big). \end{aligned} \tag{12.49d}$$

However, (12.49d) is equivalent to the following equations for the columns z_j, $j = 1, \ldots, B$,

$$\mathring{K}(x)z_j = E\left(\frac{1}{\sigma_y^U(\omega)^2}\right)b_j(x), \quad j = 1, \ldots, m. \tag{12.50}$$

With equations (12.50) for z_j, $j = 1, \ldots, m$, the expected cost function $\overline{G_1^*}(x)$ can be represented now by

$$\overline{G_1^*}(x) = tr(z_1, z_2, \ldots, z_m). \tag{12.51}$$

Having (12.50), (12.51), the deterministic substitute problems (12.35a)–(12.35d) and (12.43) can be represented as follows:

Theorem 12.2 (Expected cost-based optimization (ECBO)) *Suppose that W_{i0}, $i = 1, \ldots, B$, are given fixed weight matrices. Then the expected cost-based optimization problem (12.35a)–(12.35c) can be represented by*

$$\min \quad \overline{G}_0(x) \tag{12.52a}$$

s.t.

$$tr(z_1, z_2, \ldots, z_m) \le \Gamma_1 \tag{12.52b}$$

$$\mathring{K}(x)z_j = E\left(\frac{1}{\sigma_y^U(\omega)^2}\right)b_j(x), \quad j = 1, \ldots, m \tag{12.52c}$$

$$x \in D, \tag{12.52d}$$

where the vectors $b_j = b_j(x)$, $j = 1, \ldots, m$, are given by (12.49a).

Obviously, (12.52a)–(12.52d) is an ordinary deterministic parameter optimization problem having the additional auxiliary variables $z_j \in \mathbb{R}^m$, $j = 1, \ldots, m$. In many important cases the external generalized forces $P = P\big(a(\omega)\big)$ does not depend on the design vector x. In this case $b_1, b_2, \ldots, b_m$ are the fixed columns of the matrix $B = E\,P\big(a(\omega)\big)P\big(a(\omega)\big)^T$ of second-order moments of the random vector of external generalized forces $P = P\big(a(\omega)\big)$, see (12.49a)–(12.49c).

For the second substitute problem we get this result:

Theorem 12.3 (Minimum expected costs (MEC)) *Suppose that W_{0i}, $i = 1, \ldots, B$, are given fixed weight matrices. Then the minimum expected cost problem (12.43) can be represented by*

$$\min \quad \overline{G}_0(x) + tr(z_1, z_2, \ldots, z_m) \tag{12.53a}$$

s.t.

$$\mathring{K}(x) z_j = E\left(\frac{1}{\sigma_y^U(\omega)^2}\right) b_j(x), \quad j = 1, \ldots, m \tag{12.53b}$$

$$x \in D. \tag{12.53c}$$

Remark 12.2 According to (12.47b) and (12.46c), the matrix $\tilde{K} = \tilde{K}(x)$ is a simple function of the design vector x.

12.5.1 Symmetric, Non-uniform Yield Stresses

If a representation of

$$U(x) = EU\big(a(\omega), x\big) = EK\big(a(\omega), x\big)^{-1} = \beta(\omega)\mathring{K}(x)^{-1},$$

see (12.29d), (12.30a), (12.30b), of the type (12.47d) does not hold, then we may apply the approximative procedure described in the following.

First, the probability distribution $P_{a(\cdot)}$ of the random vector $a = a(\omega)$ of model parameters is approximated, as far it concerns the subvector $a_I = a_I(\omega)$ of $a = a(\omega)$ of model parameters arising in the matrix

$$K = K\big(a(\omega), x\big) = K\big(a_I(\omega), x\big),$$

by a discrete distribution

$$\hat{P}_{a_I(\cdot)} := \sum_{s=1}^{N} \alpha_s \varepsilon_{a_I^{(s)}} \tag{12.54}$$

having realizations $a_I^{(s)}$ taken with probabilities α_s, $s = 1, \ldots, N$.

Then, the matrix function $U = U(x)$ can be approximated by

$$\hat{U}(x) := \sum_{s=1}^{N} \alpha_s K^{(s)}(x)^{-1}, \tag{12.55a}$$

where

$$K^{(s)}(x) := K(a_I^{(s)}, x) = C K_0(a_I^{(s)}, x) C^T, \tag{12.55b}$$

see (12.30b). Consequently, the expected cost function $\overline{G_1^*} = \overline{G_1^*}(x)$ is approximated by

$$\begin{aligned} \hat{\overline{G_1^*}}(x) &:= tr\hat{U}(x) E P\big(a(\omega), x\big) P\big(a(\omega), x\big)^T \\ &= \sum_{s=1}^{N} \alpha_s tr K^{(s)}(x)^{-1} E P\big(a(\omega), x\big) P\big(a(\omega), x\big)^T. \end{aligned} \tag{12.56}$$

Corresponding to (12.49d), we now define the auxiliary matrix variables

$$\begin{aligned} z^{(s)} &= (z_1^{(s)}, z_2^{(s)}, \ldots, z_m^{(s)}) := K^{(s)}(x)^{-1} B(x) \\ &= \big(K^{(s)}(x)^{-1} b_1(x), K^{(s)}(x)^{-1} b_2(x) \ldots, K^{(s)}(x)^{-1} b_m(x)\big), \end{aligned} \tag{12.57}$$

where $B = B(x)$ is defined again by (12.49a). Thus, for the columns $z_j^{(s)}$, $j = 1, \ldots, m$, we obtain the conditions

$$K^{(s)}(x) z_j^{(s)} = b_j(x), \quad j = 1, \ldots, m, \tag{12.58}$$

for each $s = 1, \ldots, N$. According to (12.56) and (12.60), the approximate expected cost function $\hat{\overline{G_1^*}} = \hat{\overline{G_1^*}}$ reads

$$\hat{\overline{G_1^*}}(x) = \sum_{s=1}^{N} \alpha_s tr(z_1^{(s)}, z_2^{(s)}, \ldots, z_m^{(s)}), \tag{12.59}$$

where $z_j^{(s)}$, $j = 1, \ldots, m$, $s = 1, \ldots, N$, are given by (12.58).

Because of the close relationship between the representations (12.59) and (12.51) for $\hat{\overline{G_1^*}}$, $\overline{G_1^*}$, approximate mathematical optimization problems result from (12.59) which are similar to (12.52a)–(12.52d), (12.53a)–(12.53c), respectively.

12.5.2 Non Symmetric Yield Stresses

In generalization of (12.44), here we suppose

$$\sigma_{yi}^{U}(\omega) = \gamma_i^{U}\sigma_y(\omega),\ \sigma_{yi}^{L}(\omega) = \gamma_i^{L}\sigma_y(\omega), \tag{12.60}$$

where $\sigma_y = \sigma_y(\omega) > 0$ is a nonnegative random variable with a given probability distribution, and $\gamma_i^{U} > 0$, $\gamma_i^{L} < 0$, $i = 1, \dots B$, denote given, fixed yield coefficients. However, if (12.60) holds, then

$$\frac{\sigma_{yi}^{U} \pm \sigma_{yi}^{L}}{2} = \sigma_y \frac{\gamma_i^{U} \pm \gamma_i^{L}}{2} \tag{12.61a}$$

and

$$\frac{\eta_i^{U} \pm \eta_i^{L}}{2} = \sigma_y \frac{\widetilde{\eta}_i^{U}(x) \pm \widetilde{\eta}_i^{L}(x)}{2}, \tag{12.61b}$$

where, cf (12.4b, (12.4c), (12.4g), (12.4h),

$$\widetilde{\eta}_i^{\Lambda}(x) := \min\left\{\frac{|\gamma_i^{\Lambda}|A_i(x)}{N_{i0}}, \frac{\gamma_i^{U}W_{ipl}(x)}{M_{i0}}\right\}, \Lambda = L, U. \tag{12.61c}$$

Corresponding to (12.45a), (12.45b), from (12.8f), (12.8g) we obtain

$$F_i^c\Big(a(\omega), x\Big) = \sigma_y(\omega)\mathring{F}_i^c(x) \tag{12.62a}$$

$$\varrho_i^c\Big(a(\omega), x\Big) = \sigma_y(\omega)\mathring{\varrho}_i(x), \tag{12.62b}$$

where the deterministic functions

$$\mathring{F}_i = \mathring{F}_i^c(x), \quad \mathring{\varrho}_i = \mathring{\varrho}_i(x) \tag{12.62c}$$

follow from (12.8f), (12.8g), resp., by inserting formula (12.60) and extracting then the random variable $\sigma(\omega)$. Because of (12.62a)–(12.62c), the generalized stiffness matrix $K = K(a, x)$ can be represented again in the form (12.47a), hence,

$$K\Big(a(\omega), x\Big) = \sigma_y^2(\omega)\mathring{K}(x), \tag{12.63a}$$

where the deterministic matrix function $\mathring{K} = \mathring{K}(x)$ is defined corresponding to (12.47b). Furthermore, according to (12.21h) and (12.62a), for $F^c\Big(a(\omega), x\Big)$ we have

$$F^c\Big(a(\omega), x\Big) = \sigma(\omega)\mathring{F}^c(x), \tag{12.63b}$$

where

$$\mathring{F}^c(x) := \left(\mathring{F}_1^c(x)^T, \ldots, \mathring{F}_B^c(x)^T\right)^T. \tag{12.63c}$$

Thus, due to (12.32b), (12.32c), for the vector $F^{c*} = F^{c*}\Big(a(\omega), x\Big)$ defined by (12.32a)–(12.32c) we find

$$F^{c*}\Big(a(\omega), x\Big) = \sigma(\omega) F^{c**}(x) \tag{12.63d}$$

with

$$F^{c**}(x) := (H^T H)^{-1} H^T \mathring{F}^c(x). \tag{12.63e}$$

Inserting now (12.63a), (12.63d) into formula (12.34b), for the (approximate) minimum total costs we finally have, cf. (12.47c), (12.48),

$$\begin{aligned} G_1^{a*}\Big(a(\omega), x\Big) &= \frac{1}{\sigma(\omega)^2} tr\mathring{K}(x)^{-1}\left(\sigma(\omega) C F^{c**}(x) - P\Big(a(\omega), x\Big)\right) \\ &\qquad \times \left(\sigma(\omega) C F^{c**}(x) - P\Big(a(\omega), x\Big)\right)^T \\ &= \frac{1}{\sigma(\omega)^2} tr\mathring{K}(x)^{-1} P\Big(a(\omega), x\Big) P\Big(a(\omega), x\Big)^T \\ &\quad - \frac{1}{\sigma(\omega)} tr\mathring{K}(x)^{-1}\left(C F^{c**}(x) P\Big(a(\omega), x\Big)^T\right. \\ &\quad \left. + P\Big(a(\omega), x\Big) F^{c**}(x)^T C^T\right) + tr\mathring{K}(x)^{-1} C F^{c**}(x) F^{c**}(x)^T C^T. \end{aligned} \tag{12.64}$$

The minimum expected cost function $\overline{G_1^{a*}}(x)$ is then given, cf. (12.48), by

$$\begin{aligned} \overline{G_1^{a*}}(x) &= E\left(\frac{1}{\sigma(\omega)^2}\right) tr\mathring{K}(x)^{-1} B(x) \\ &\quad - E\left(\frac{1}{\sigma(\omega)}\right) tr\mathring{K}(x)^{-1}\left(C F^{c**}(x)\overline{P}(x)^T + \overline{P}(x) F^{c**}(x)^T C^T\right) \\ &\quad + tr\mathring{K}(x)^{-1} C F^{c**}(x) F^{c**}(x)^T C^T, \end{aligned} \tag{12.65}$$

where $\overline{P} = \overline{P}(x)$, $B = B(x)$ are again given by (12.49a), (12.49b). Obviously, also in the present more general case $\overline{G_1^{a*}}$ can be represented by

$$\overline{G_1^{a*}}(x) = tr Z(x), \tag{12.66a}$$

where the matrix $Z = Z(x)$ is given by

$$Z(x) := \mathring{K}(x)^{-1} \left(E\left(\frac{1}{\sigma_y(\omega)^2}\right) B(x) - E\left(\frac{1}{\sigma_y(\omega)}\right) \right.$$
$$\left. \times \left(CF^{c**}(x)\overline{P}(x)^T + \overline{P}(x)F^{c**}(x)^T C^T \right) + CF^{c**}(x)F^{c**}(x)^T C^T \right), \tag{12.66b}$$

see (12.51). Hence, due to the close relationship between (12.49d), (12.51) and (12.66a), (12.66b), the deterministic substitute problems stated in Sect. 12.4 can be treated as described in Sect. 12.5.1.

12.6 Reliability Analysis

For the approximate computation of the probability of survival $p_s = p_s(x)$ in Sect. 12.4.1 a first method was presented based on certain probability inequalities. In the following subsection we suppose that $x = x_0$ is a fixed design vector, and the vector of yield stresses

$$\sigma_y = \begin{pmatrix} (\sigma_{yi}^L)_{i=1,\ldots,B} \\ (\sigma_{yi}^U)_{i=1,\ldots,B} \end{pmatrix} = \sigma_{y0}$$

is a given deterministic vector of material strength parameters. Moreover, we assume that the weight matrix W_0, cf. (12.20c), (12.21a)–(12.21c), is fixed. According to (12.8f), (12.8g), (12.21h) and (12.30a), (12.30b), the vectors $\widetilde{F}^c$, $\widetilde{Q}$ and the generalized stiffness matrix $K = K(\sigma_{y0}, x_0)$ are given, fixed quantities. Hence, in this case the cost function

$$G_1^*(a, x) = g_1^*(P) := (C\widetilde{F}^c - P)^T K^{-1}(C\widetilde{F}^c - P), \tag{12.67}$$

see (12.28g), is a quadratic, strictly convex function of the m-vector P of external generalized forces. Hence, the condition $G_1^*(a, x_0) \le g_1$, see (12.37a), (12.37b), or

$$g_1^*(P) \le g_1 \tag{12.68a}$$

describes an ellipsoid in the space $\mathbb{R}^m$ of generalized forces.

In case of normal distributed external generalized forces $P = P(\omega)$, the probability

$$p_s(x_0; g_1) := P\Big(G_1^*\Big(\sigma_{y0}, P(\omega), x_0\Big) \le g_1 \Big)$$
$$= P\Big(g_1^*\Big(P(\omega)\Big) \le g_1 \Big) \tag{12.68b}$$

can be determined approximatively by means of linearization

$$g_1^*(P) = g_1^*(P^*) + \nabla_P g_1^*(P^*)^T (P - P^*) + \cdots \tag{12.69}$$

at a so-called design point P^*, see [2, 3, 9]. Since $g_1^* = g_1^*(P)$ is convex, we have

$$g_1^*(P) \geq g_1^*(P^*) + \nabla_P g_1^*(P^*)^T (P - P^*), \;\; P \in \mathbb{R}^m, \tag{12.70a}$$

and $p_s(x_0; g_1)$ can be approximated *from above* by

$$\tilde{p}_s(x_0; g_1) := P\Big(g_1^*(P^*) + \nabla_P g_1^*(P^*)^T \Big(P(\omega) - P^*\Big) \leq g_1\Big) \tag{12.70b}$$

$$= P\Big(\nabla_P g_1^*(P^*)^T P(\omega) \leq g_1 - g_1^*(P^*) + \nabla_P g_1^*(P^*)^T P^*\Big)$$

$$= \Phi\left(\frac{g_1 - g_1^*(P^*) + \nabla_P g_1^*(P^*)^T (P^* - \overline{P})}{\sqrt{\nabla_P g_1^*(P^*)^T \mathrm{cov}\Big(P(\cdot)\Big) \nabla_P g_1^*(P^*)}}\right), \tag{12.70c}$$

where Φ denotes the distribution function of the $N(0, 1)$-normal distribution, $\mathrm{cov}\Big(P(\cdot)\Big)$ denotes the covariance matrix of $P = P(\omega)$, and the gradient $\nabla_P g_1^*(P^*)$ reads

$$\nabla_P g_1^*(P^*) = -2K^{-1}(C\widetilde{F}^c - P^*). \tag{12.70d}$$

Moreover,

$$\overline{P} := E\,P(\omega) \tag{12.70e}$$

denotes the mean of the external vector of generalized forces $P = P(\omega)$. In practice, the following two cases are taken into account [2, 3, 9]:

Case 1: Linearization at $P^* := \overline{P}$
Under the above assumptions in this case we have

$$\tilde{p}_s(x_0; g_1) = \Phi\left(\frac{g_1 - g_1^*(\overline{P})}{\sqrt{\nabla_P g_1^*(\overline{P})^T \mathrm{cov}\Big(P(\cdot)\Big) \nabla_P g_1^*(\overline{P})}}\right) \geq p_s(x_0; g_1). \tag{12.71}$$

Case 2: Linearization at a boundary point P^* of $[g_1^*(P) \leq g_1]$
Here it is $g_1(P^*) = g_1$ and therefore

$$\tilde{p}_s(x_0; g_1) = \Phi\left(\frac{\nabla_P g_1^*(P^*)^T (P^* - \overline{P})}{\sqrt{\nabla_P g_1^*(P^*)^T \operatorname{cov}\Big(P(\cdot)\Big)\nabla_P g_1^*(P^*)}}\right). \tag{12.72}$$

Because of (12.70a), for each boundary point P^* we have again

$$p_s(x_0; g_1) \le \tilde{p}_s(x_0; g_1). \tag{12.73}$$

Boundary points P^* of the ellipsoid $\Big[g_1^*(P) \le g_1\Big]$ can be determined by minimizing a linear form $c^T P$ on $\Big[g_1^*(P) \le g_1\Big]$. Thus, we consider [7] the convex minimization problem

$$\min_{(C\widetilde{F}^c - P)^T K^{-1} (C\widetilde{F}^c - P) \le g_1} c^T P, \tag{12.74}$$

where c is a given, fixed m-vector.

By means of Lagrange techniques we obtain this result.

Theorem 12.4 *For each vector $c \neq 0$ the unique minimum point of (12.74) reads*

$$P^* = C\widetilde{F}^c - \sqrt{\frac{g_1}{c^T K c}}\, K c. \tag{12.75a}$$

The gradient of $g_1^* = g_1^*(P)$ at P^* is then given by

$$\nabla_P g_1^*(P^*) = -2\sqrt{\frac{g_1}{c^T K c}}\, c. \tag{12.75b}$$

Consequently, for the quotient q arising in formula (12.72) we get

$$q = q(c) := \frac{\sqrt{g_1}\sqrt{c^T K c} - c^T (C\widetilde{F}^c - \overline{P})}{\sqrt{c^T \operatorname{cov}\Big(P(\cdot)\Big) c}}. \tag{12.76a}$$

Obviously, this function fulfills the equation

$$q(\lambda c) = q(c),\ \lambda > 0, \tag{12.76b}$$

for each m-vector c such that $c^T \operatorname{cov}\Big(P(\cdot)\Big) c \neq 0$.

Since

$$\tilde{p}_s(x_0; g_1) = \Phi(q(c)) \ge p_s(x_0; g_1),\ c \neq 0, \tag{12.77a}$$

see (12.72), (12.73), the best upper bound $\tilde{p}_s(x_0, g_1)$ can be obtained by solving the minimization problem

$$\min_{c \neq 0} q(c). \tag{12.77b}$$

Because of (12.76b), problem (12.77b) is equivalent to the convex optimization problem

$$\min_{c^T cov\left(P(\cdot)\right)c=1} \sqrt{g_1}\sqrt{c^T K c} - c^T (C\widetilde{F}^C - \overline{P}), \tag{12.78}$$

provided that $\mathrm{cov}(P(\cdot))$ is regular. Representing the covariance matrix of $P = P(\omega)$ by

$$\mathrm{cov}\Big(P(\cdot)\Big) = Q^T Q$$

with a regular matrix Q, problem (12.78) can be represented also by

$$\min_{\|w\|=1} \sqrt{g_1}\sqrt{w^T Q^{-1T} K Q^{-1} w} - w^T Q^{-1T} (C\widetilde{F}^c - \overline{P}). \tag{12.79}$$

12.7 Numerical Example: 12-Bar Truss

The new approach for the optimal design of elastoplastic mechanical structures under stochastic uncertainty is illustrated now by means of the 12-bar truss according to Fig. 12.2.

Suppose that $L = 1000\,\mathrm{mm}$, $E = 7200\dfrac{N}{mm^2}$ is the elasticity modulus, and the yield stresses with respect to tension and compression are given by $\sigma_y^U = -\sigma_y^L = \sigma_y = 216\dfrac{N}{mm^2}$. Furthermore, assume that the structure is loaded by the deterministic force components

$$P_{1x} = P_{3y} = 10^5\, N$$

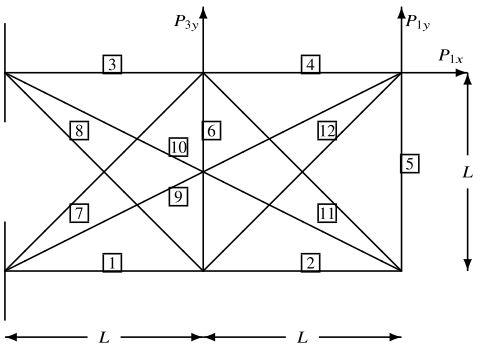


Fig. 12.2 12-bar truss

and the random force component

$$P_{1y} \cong N(\mu, \sigma^2)$$

having a normal distribution with mean μ and variance σ^2. The standard deviation σ is always 10% of the mean μ.

The numerical results presented in this section have been obtained by Dipl.Math.oec. Simone Zier, Institute for Mathematics and Computer Applications, Federal Armed Forces University, Munich.

The equilibrium matrix C of the 12-bar truss is given by

$$C = \begin{pmatrix} 0 & 0 & 0 & 1.0 & 0 & 0 & 0 & 0 & 0.894427 & 0 & 0 & 0.707107 \\ 0 & 0 & 0 & 0 & 1.0 & 0 & 0 & 0 & 0.447214 & 0 & 0 & 0.707107 \\ 0 & 1.0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.894427 & 0.707107 & 0 \\ 0 & 0 & 0 & 0 & -1.0 & 0 & 0 & 0 & 0 & -0.447214 & -0.707107 & 0 \\ 0 & 0 & 1.0 & -1.0 & 0 & 0 & 0.707107 & 0 & 0 & 0 & -0.707107 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.0 & 0.707107 & 0 & 0 & 0 & 0.707107 & 0 \\ 1.0 & -1.0 & 0 & 0 & 0 & 0 & 0 & 0.707107 & 0 & 0 & 0 & -0.707107 \\ 0 & 0 & 0 & 0 & 0 & -1.0 & 0 & -0.707107 & 0 & 0 & 0 & -0.707107 \end{pmatrix}. \tag{12.80}$$

Note that under the above assumptions, condition (12.22b) holds.

Since in the present case of a truss we have $H = I$ ($B \times B$ identity matrix), cf. (12.5b) and (12.21h), the matrix $K_0 = K_0(a, x) = K_0(\sigma_y, x)$, see (12.30b) and (12.46b), is a diagonal matrix represented by

$$K_0(\sigma, x) = \operatorname{diag}\left(\left(\frac{\varrho_i(x)}{w_i^0}\right)^2\right), \tag{12.81a}$$

cf. (12.30d). Here, w_i^0 is the element of the 1×1 weight matrix W_{i0}, and $\varrho_i = \varrho_i(x)$ is defined, cf. (12.7b), by

$$\varrho_i(x) = \frac{F_i^U - F_i^L}{2} = \sigma_y A_i(x), \quad i = 1, \ldots, B. \tag{12.81b}$$

Defining

$$\tilde{w}_i = \tilde{w}_i(x) := \left(\frac{\varrho_i(x)}{w_i^0}\right)^2, \quad i = 1, \ldots, B, \tag{12.81c}$$

the generalized stiffness matrix $K = K(\sigma_y, x)$, see (12.30a), reads

$$K(\sigma_y, x) = C K_0(\sigma_y, x) C^T$$

$$= \begin{pmatrix}
\tilde{w}_4 + 0.8\tilde{w}_9 + 0.5\tilde{w}_{12} & 0.4\tilde{w}_9 + 0.5\tilde{w}_{12} & 0 & 0 \\
0.4\tilde{w}_9 + 0.5\tilde{w}_{12} & \tilde{w}_5 + 0.2\tilde{w}_9 + 0.5\tilde{w}_{12} & 0 & -\tilde{w}_5 \\
0 & 0 & \tilde{w}_2 + 0.8\tilde{w}_{10} + 0.5\tilde{w}_{11} & 0.4\tilde{w}_{10} - 0.5\tilde{w}_{11} \\
0 & -\tilde{w}_5 & -0.4\tilde{w}_{10} - 0.5\tilde{w}_{11} & \tilde{w}_5 + 0.2\tilde{w}_{10} + 0.5\tilde{w}_{11} \\
-\tilde{w}_4 & 0 & -0.5\tilde{w}_{11} & 0.5\tilde{w}_{11} \\
0 & 0 & 0.5\tilde{w}_{11} & -0.5\tilde{w}_{11} \\
-0.5\tilde{w}_{12} & -0.5\tilde{w}_{12} & -\tilde{w}_2 & 0 \\
-0.5\tilde{w}_{12} & -0.5\tilde{w}_{12} & 0 & 0
\end{pmatrix}$$

$$\begin{pmatrix}
-\tilde{w}_4 & 0 & -0.5\tilde{w}_{12} & -0.5\tilde{w}_{12} \\
0 & 0 & -0.5\tilde{w}_{12} & -0.5\tilde{w}_5 \\
-0.5\tilde{w}_{11} & 0.5\tilde{w}_{11} & -\tilde{w}_2 & 0 \\
0.5\tilde{w}_{11} & -0.5\tilde{w}_{11} & 0 & 0 \\
\tilde{w}_3 + \tilde{w}_4 + 0.5\tilde{w}_7 + 0.5\tilde{w}_{11} & 0.5\tilde{w}_7 - 0.5\tilde{w}_{11} & 0 & 0 \\
0.5\tilde{w}_7 - 0.5\tilde{w}_{11} & \tilde{w}_6 + 0.5\tilde{w}_7 + 0.5\tilde{w}_{11} & 0 & -\tilde{w}_6 \\
0 & 0 & \tilde{w}_1 + \tilde{w}_2 + 0.5\tilde{w}_8 + 0.5\tilde{w}_{12} & -0.5\tilde{w}_8 + 0.5\tilde{w}_{12} \\
0 & -\tilde{w}_6 & -0.5\tilde{w}_8 + 0.5\tilde{w}_{12} & \tilde{w}_6 + 0.5\tilde{w}_8 + 0.5\tilde{w}_{12}
\end{pmatrix}. \tag{12.82}$$

12.7.1 Numerical Results: MEC

In the present case the cost factors γ_{i0} in the primary cost function $G_0(x) = \overline{G_0}(x)$, cf. (12.1a), are defined by

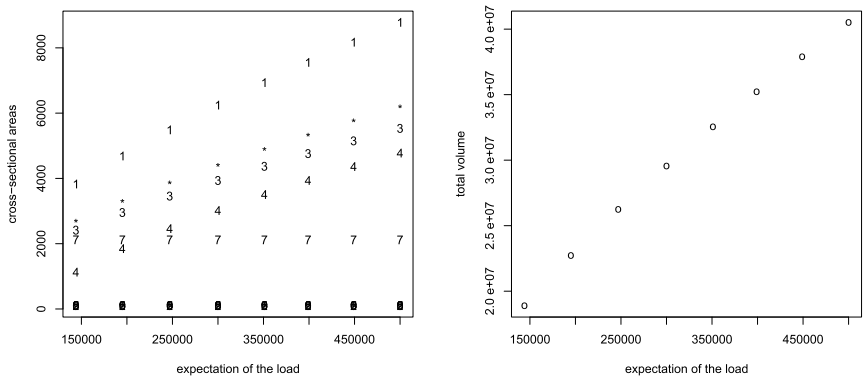
$$\gamma_{i0} := \frac{1}{V_0} = 6,4 \cdot 10^{-4} \left[\frac{1}{mm^3} \right]$$

corresponding to the chosen reference volume $V_0 = 1562,5\,\mathrm{mm}^3$. Thus, the cost function $\overline{G_0}(x)$ and the recourse cost function $\overline{G_1^*}(x)$ are dimensionless, cf. Sect. 4.2. Furthermore, the weight factors in the recourse costs $G_1(x)$ are defined by

$$w_i^0 = 100.$$

In Fig. 12.3a, b the optimal cross-sectional areas A_i^*, $i = 1, \ldots, 12$, and the total volume are shown as functions fo the expectation $\overline{P}_{1y} = E P_{1y}(\omega)$ of the random force component $P_{1y} = P_{1y}(\omega)$. With increasing expected force $\overline{P}_{1y}$, the cross-sectional areas of bar 1,3,4,8,12 are increasing too, while the other remain constant or are near zero. The resulting optimal design of the truss can be seen in Fig. 12.4. Here, bars with cross-sectional areas below $A_{\min} = 100\,\mathrm{mm}^2$ have been deleted.

In Figs. 12.3a and 12.5a by the symbol "$*$" the almost equal optimal cross-sectional areas of bar 8 and 12 are marked.



(a) Optimal cross-sectional areas (b) Total volume

Fig. 12.3 Optimal design using (MEC)

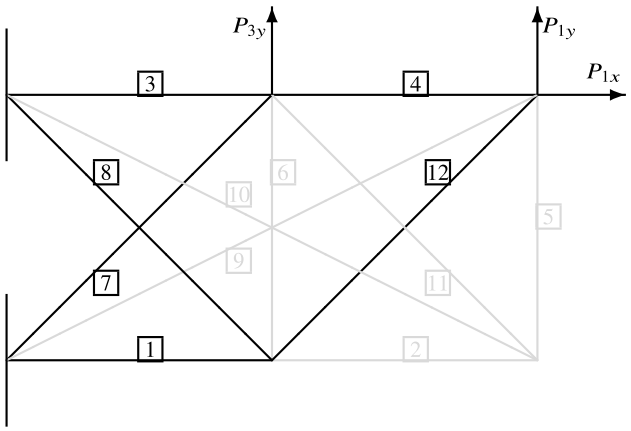
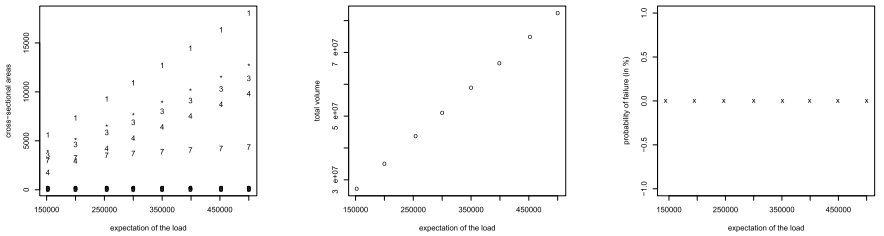


Fig. 12.4 Optimal 6-bar truss using MEC

The probability of failure of an (MEC)-optimal truss is always zero showing also the robustness of the optimal 6-bar truss according to Fig. 12.4.

12.7.2 Numerical Results: ECBO

The related numerical results obtained for the expected cost-based optimization problem (ECBO) are presented in Fig. 12.5a–c. Here, the optimal cross-sectional areas, the expected minimum volume, and the related probability of failure are represented again as functions of the expected form $\overline{P}_{1y}$. The resulting optimal design is the same



(a) Optimal cross-sectional areas (b) Expected minimum volume $\overline{G}_0$ (c) Probability of failure

Fig. 12.5 Optimal design using (ECBO)

as in (MEC), where in this case the probability of failure is also zero, which confirms again the robustness of this optimal design.

References

1. Bullen, P.: Handbook of Means and Their Inequalities. Kluwer Academic Publishing, Dordrecht (2003)
2. Ditlevsen, O., Madsen, H.: Structural Reliability Methods. Wiley, New York (1996)
3. Frangopol, D.: Reliability-based optimum structural design. In: Sundararajan, C. (ed.) Probabilistic Structural Mechanics Handbook, pp. 352–387. Chapman & Hall, New York (1995)
4. Hardy, G., Littlewood, J., Pólya, G.: Inequalities. Cambridge University Press, London (1973)
5. Kemenjarzh, J.: Limit Analysis of Solids and Structures. CRC Press, Boca Raton (1996)
6. Marti, K.: Stochastic optimization methods in optimal engineering design under stochastic uncertainty. ZAMM **83**(11), 1–18 (2003)
7. Marti, K.: Reliability analysis of technical systems structures by means of polyhedral approximation of the safe unsafe domain. GAMM Mitteilungen **30**(2), 211–254 (2007)
8. Marti, K.: Stochastic Optimization Methods, 2nd edn. Springer, Berlin (2008). https://doi.org/10.1007/978-3-540-79458-5
9. Schuëller, G., Gasser, M.: Some basic principles of reliability-based optimization (rbo) of structure and mechanical components. In: Marti, K., Kall, P. (eds.) Stochastic Programming Methods and Technical Applications. Lecture Notes in Economics and Mathematical Systems (LNEMS), vol. 458, pp. 80–103. Springer, Berlin (1998)
10. Spillers, W.: Automated Structural Analysis: An Introduction. Pergamon, New York (1972)

Chapter 13
Maximum Entropy Techniques

Abstract Finally, in this chapter the inference and decision strategies applied in stochastic optimization methods are considered in more detail: A large number of optimization problems arising in engineering, control, and economics can be described by the minimization of a certain (cost) function $v = v(a, x)$ depending on a random parameter vector $a = a(\omega)$ and a decision vector $x \in D$ lying in a given set D of feasible decision, design or control variables. Hence, in order to get *robust optimal decisions*, i.e., optimal decisions being most insensitive with respect to variations of the random parameter vector $a = a(\omega)$, the original optimization problem is replaced by the deterministic substitute problem which consists in the minimization of the expected objective function $\mathbf{E}v = \mathbf{E}v(a(\omega), x)$ subject to $x \in D$. Since the true probability distribution λ of $a = a(\omega)$ is not exactly known in practice, one has to replace λ by a certain estimate or guess β. Consequently, one has the following *inference and decision problem*:

- *inference/estimation step*
  Determine an estimation β of the true probability distribution λ of $a = a(\omega)$,
- *decision step*
  Determine an optimal solution x^* of $min \int v(a(\omega), x)\beta(d\omega)$ s.t. $x \in D$.

Computing approximation, estimation β of λ, the criterion for judging an approximation β of λ should be based on its utility for the decision-making process, i.e., one should weight the approximation error according to its influence on decision errors, and the decision errors should be weighted in turn according to the loss caused by an incorrect decision.
Based on inferential ideas developed among others by Kerridge, Kullback, in this chapter generalized decision-oriented inaccuracy and divergence functions for probability distributions λ, β are developed, taking into account that the outcome β of the inferential stage is used in a subsequent (ultimate) decision-making problem modeled by the above-mentioned stochastic optimization problem. In addition, *stability properties* of the inference and decision process

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9_13

$$\lambda \longrightarrow \beta \longrightarrow x \in D_\epsilon(\beta)$$

are studied, where $D_\epsilon(\beta)$ denotes the set of $\epsilon-$optimal decisions with respect to probability distribution $P_{a(\cdot)} = \beta$ of the random parameter vector $a = a(\omega)$.

13.1 Uncertainty Functions Based on Decision Problems

13.1.1 Optimal Decisions Based on the Two-Stage Hypothesis Finding (Estimation) and Decision-Making Procedure

According to the considerations in the former chapters, in the following we suppose that $v = v(\omega, x)$ denotes the costs or the loss arising in a decision or design problem if the action or design $x \in D$ is taken, and the elementary event $\tilde{\omega} = \omega$ has been realized. Note that $v = v(\omega, x)$ is an abbreviation for $v = v(a(\omega), x)$, where $a(\omega)$ denotes the vector of all random parameters under consideration. As a deterministic substitute for the optimal decision/design problem under stochastic uncertainty

$$\text{minimize } v(\omega, x) \quad \text{s.t. } x \in D \tag{13.1}$$

we consider, cf. Chap. 1, the expectation or mean value minimization problem

$$\text{minimize } v(\lambda, x) \quad \text{s.t. } x \in D, \tag{13.2a}$$

where

$$v(\lambda, x) = v(P_{\tilde{\omega}}, x) := Ev(\tilde{\omega}, x) = \int v(\omega, x)\lambda(d\omega). \tag{13.2b}$$

Here, "E" denotes the expectation operator, and the probability distribution λ on the measurable space $(\Omega, \mathfrak{A})$ denotes the true probability distribution $P_{\tilde{\omega}} := \lambda$ of the random element ω. We may assume that λ lies in a certain given set Λ of probability measures on $\mathfrak{A}$ (a priori information about λ). In the following we suppose that all integrals, expectations, probabilities, resp., under consideration exist and are finite.

Because in practice also the true probability distribution "$P_{\tilde{\omega}} = \lambda$" of ω is not known in general, one works mostly with the following *two-step Inference and Decision Procedure (IDP)*, according to Fig. 13.1:

Step I: Accept the hypothesis "$P_{\tilde{\omega}} = \beta$". Hence, work with the hypothesis that $\tilde{\omega}$ has the distribution β, where β results from a certain estimation or hypothesis-finding procedure (suitable to (Λ, D, v));

Step II: Instead of (1), solve the approximate optimization problem:

$$\text{minimize } v(\beta, x) \quad \text{s.t. } x \in D \tag{13.3}$$

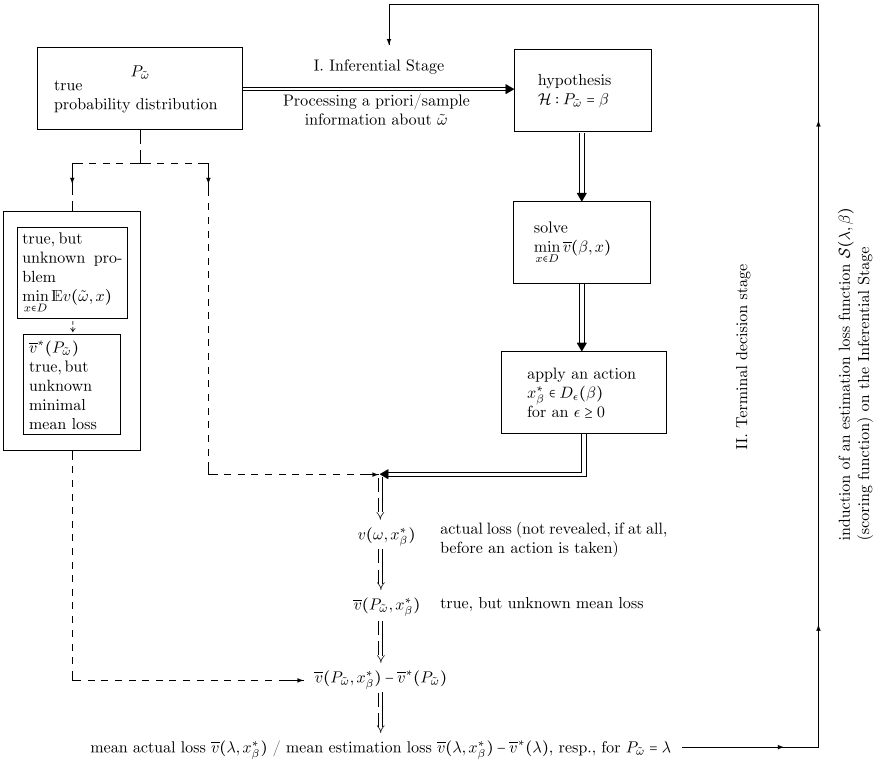


Fig. 13.1 Inference and Decision Procedure (IDP)

and take an $\varepsilon-$optimal decision $x \in D_\varepsilon(\beta)$ with an appropriate bound $\varepsilon > 0$. Here, the set $D_\varepsilon(\beta)$ of $\varepsilon-$optimal decisions is defined by

$$D_\varepsilon(\beta) := \{x \in D : v(\beta, x) \leq v^*(\beta) + \varepsilon\}. \tag{13.4a}$$

Of course,

$$D_0(\beta) := \{x \in D : v(\beta, x) = v^*(\beta)\} \tag{13.4b}$$

denotes the set of optimal solutions of (13.3). Note that in (13.4a), (13.4b) we use the minimum value of (13.3):

$$v^*(\beta) := \inf\{v(\beta, x) : x \in D\}. \tag{13.5}$$

Remark 13.1 Hypothesis-finding in case that there is some a priori information, but **no** sample information $\omega^N = (\omega_1, ..., \omega_N)$ of $\tilde{\omega}$.

In the above case a so-called $e-$estimate β of $P_{\tilde{\omega}}$ can be applied which is defined as follows:

Definition 13.1 Let $e : \Lambda \times \Lambda \to \mathbb{R}$ denote a function on the product $\Lambda \times \Lambda$ of the given set Λ of probability measures—containing the true distribution λ of $\tilde{\omega}$—such that $e(\lambda, \pi)$ can be considered as a measure for the error selecting the distribution π, while $P_{\tilde{\omega}} = \lambda$ is the true distribution. An *$e-$estimate of* $P_{\tilde{\omega}}$ is then a distribution $\beta \in \Lambda$ such that

$$\sup_{\lambda \in \Lambda} e(\lambda, \beta) \; = \inf_{\pi \in \Lambda} \; (\sup_{\lambda \in \Lambda} \; e(\lambda, \pi) \;). \tag{13.6}$$

If $e(\cdot, \cdot)$ is a metric on Λ, then the $e-$estimate β of $P_{\tilde{\omega}}$ is also called a "*Tchebychev-center*" of Λ.

Though in many cases the approximation, estimation β of λ is computed by standard statistical estimation procedures, the criterion for judging an approximation β of λ should be based on its utility for the decision-making process, i.e., one should weight the approximation error according to its influence on decision errors, and the decision errors should be weighted in turn according to the loss caused by an incorrect decision, cf. [12, 15]. A detailed consideration of this concept is given in the following.

In order to study first the properties of the above defined 2-step procedure **(I, II)**, resulting from using an estimation/approximation of the unknown or only partly known parameter distribution, we suppose that the set $D_0(\beta)$ of optimal decisions with respect to β, see (13.4b), is non empty. Because $P_{\tilde{\omega}} = \lambda$ is the true distribution, according to the 2-step procedure (I, II) we (I), replacing λ by its estimate β, and (II) applying then a certain β-optimal decision $x_\beta \in D_0(\beta)$, we have the expected loss $v(\lambda, x_\beta)$. Consequently,

$$H_0(\lambda, \beta) = \sup\{v(\lambda, x) : x \in D_0(\beta)\} \tag{13.7}$$

denotes therefore the maximum expected loss if $P_{\tilde{\omega}} = \lambda$ is the true distribution of $\tilde{\omega}$, and the decision maker uses a certain β-optimal decision. Because of $v(\lambda, x) \geq v^*(\lambda)$, $x \in D$, cf. (13.5), we have

$$H_0(\lambda, \beta) \geq v^*(\lambda) \tag{13.8a}$$

and

$$v^*(\lambda) = H_0(\lambda, \lambda) \tag{13.8b}$$

provided that also $D_0(\lambda) \neq \emptyset$. If $D_0(\beta) = \{x_\beta^*\}$, then

$$H_0(\lambda, \beta) = v(\lambda, x_\beta^*). \tag{13.9}$$

In case $D_0(\lambda) \neq \emptyset$, the difference

$$I_0(\lambda, \beta) := H_0(\lambda, \beta) - v^*(\lambda) = H_0(\lambda, \beta) - H_0(\lambda, \lambda) \tag{13.10}$$

is the maximum error relative to the decision-making problem (Ω, D, v), if any β-optimal decision is applied, while $P_{\tilde{\omega}} = \lambda$ is the true distribution of $\tilde{\omega}$. Obviously,

$$I_0(\lambda, \beta) \geq 0 \text{ and } I_0(\lambda, \beta) = 0 \text{ if } \beta = \lambda. \tag{13.11}$$

Though, according to the above assumption, problem (13.3) is solvable in principle, due to the complexity of mean value minimization problems, we have to confine in general with a certain $\varepsilon-$optimal solution, hence, with an element of $D_\varepsilon(\beta), \varepsilon > 0$. However, applying any decision x_β^ε of $D_\varepsilon(\beta)$, cf. (13.4a), we have to face the maximum expected loss

$$H_\varepsilon(\lambda, \beta) = \sup\{v(\lambda, x) : x \in D_\varepsilon(\beta)\}. \tag{13.12}$$

In order to study the function

$$\varepsilon \to H_\varepsilon(\lambda, \beta), \varepsilon > (=)0,$$

we introduce still the following notation:

Definition 13.2 Let V denote the **loss set** defined by

$$V := \{v(\cdot, x) : x \in D\}. \tag{13.13a}$$

Moreover, corresponding to $D_\varepsilon(\beta)$, the subset $V_\varepsilon(\beta)$ of V is defined by

$$V_\varepsilon(\beta) := \{v(\cdot, x) : x \in D_\varepsilon(\beta)\}. \tag{13.13b}$$

Based on the loss set V, the functions $H = H_\varepsilon(\lambda, \beta)$ and $v^*(\lambda)$ can be represented also as follows:

$$H_\varepsilon(\lambda, \beta) = \sup\left\{\int v(\omega)\lambda(d\omega) : v \in V_\varepsilon(\beta)\right\}, \tag{13.13c}$$

$$v^*(\lambda) := \inf\left\{\int v(\omega)\lambda(d\omega) : v \in V\right\}. \tag{13.13d}$$

Remark 13.2 According to the above assumptions, the loss set V lies in the space $L_1(\Omega, \mathfrak{A}, \pi)$ of all $\pi-$integrable functions $f = f(\omega$ for each probability distribution $\pi = \lambda, \pi = \beta$ under consideration.

Remark 13.3 Identifying the decision vector $x \in D$ with the related loss function $v = v(\omega, x)$, the set D of decisions can be identified with the related loss set V. Hence, we can consider the loss set V as the *generalized admissible set of decision*

or design vectors. On the other hand, the optimal decision problem under stochastic uncertainty can also be described by the set Ω of elementary events and a certain set V of measurable real functions $f = f(\omega)$ on Ω playing the role of loss functions related to the decision vector $x \equiv f(\cdot)$.

We have then the following properties:

Lemma 13.1 *Suppose that $D_\varepsilon(\beta) \neq \emptyset$ for all $\varepsilon > 0$.*

(I) $\varepsilon \to H_\varepsilon(\lambda, \beta)$ *is monotonous increasing on* $(0, +\infty)$;
(II) $H_\varepsilon(\lambda, \beta) \geq v^*(\lambda), \varepsilon > 0$;
(III) If the loss set V is convex, then $\varepsilon \to H_\varepsilon(\lambda, \beta)$ is concave;
(IV) If $v(\beta, x) \leq v^*(\beta) + \bar{\varepsilon}$ for all $x \in D$ and a fixed $\bar{\varepsilon} > 0$, then $H_\varepsilon(\lambda, \beta) = sup\{v(\lambda, x) : x \in D\}, \varepsilon \geq \bar{\varepsilon}$
(V) The assertions (1)–(4) hold also for $\varepsilon \geq 0$, provided that $D_0(\beta) \neq \emptyset$.

Proof Because of $D_\varepsilon(\beta) \neq \emptyset, \varepsilon > 0$, the maximum expected loss $H_\varepsilon(\lambda, \beta)$ is defined for all $\varepsilon > 0$. (I) The monotonicity of $\varepsilon \to H_\varepsilon(\lambda, \beta)$ follows from $D_\varepsilon(\beta) \subset D_\delta(\beta)$, if $\varepsilon < \delta$. (II) The inequality $v(\lambda, x) \geq v^*(\lambda), x \in D$, yields $H_\varepsilon(\lambda, \beta) = \sup\{(\lambda, x) : x \in D_\varepsilon(\beta)\} \geq v^*(\lambda)$. (III) Let be $\epsilon_1 > 0, \epsilon_2 > 0, 0 \leq \alpha \leq 1$ and $x_1 \in D_{\varepsilon_1}(\beta), x_2 \in D_{\varepsilon_2}(\beta)$. Because of the convexity of the loss set V, there exists $x_3 \in D$, such that $v(\cdot, x_3) = \alpha v(\cdot, x_1) + (1 - \alpha)v(\cdot, x_2)$. This yields then $v(\beta, x_3) = \alpha v(\beta, x_1) + (1 - \alpha)v(\beta, x_2) \leq \alpha(v^*(\beta) + \varepsilon_1) + (1 - \alpha)(v^*(\beta) + \varepsilon_2) = v^*(\beta) + \bar{\varepsilon}$ with $\bar{\varepsilon} = \alpha\varepsilon_1 + (1 - \alpha)\varepsilon_2$. Hence, $x_3 \in D_{\bar{\varepsilon}}(\beta)$ and therefore $H_{\bar{\varepsilon}}(\lambda, \beta) \geq v(\lambda, x_3) = \alpha v(\lambda, x_1) + (1 - \alpha)v(\lambda, x_2)$. Since x_1, x_2 were chosen arbitrarily, we get now $H_{\bar{\varepsilon}}(\lambda, \beta) \geq \alpha H_{\varepsilon_1}(\lambda, \beta) + (1 - \alpha)H_{\varepsilon_2}(\lambda, \beta)$. The rest of the assertion is clear. □

Remark 13.4 According to Lemma 13.1(V) we have $H_\varepsilon(\lambda, \beta) \geq H_0(\lambda, \beta), \varepsilon \geq 0$, provided that $D_0(\beta) \neq \emptyset$.

By the above result the limit "$\lim_{\varepsilon \downarrow 0}$" exists, and we have

$$H(\lambda, \beta) := \lim_{\varepsilon \downarrow 0} H_\varepsilon(\lambda, \beta) = \inf_{\varepsilon > 0} H_\varepsilon(\lambda, \beta). \tag{13.14a}$$

$$H(\lambda, \beta) \geq v^*(\lambda) \tag{13.14b}$$

and

$$H(\lambda, \beta) \geq H_0(\lambda, \beta) \geq v^*(\lambda) \quad \text{if } D_0(\beta) \neq \emptyset. \tag{13.14c}$$

A detailed study of $H(\lambda, \beta)$ and $I(\lambda, \beta) := H(\lambda, \beta) - v^*(\lambda)$ follows in Sects. 13.1 and 13.2, where we find a close relationship of H, I, resp., with the inaccuracy function of Kerridge [8], the divergence of Kullback [9]. Thus, we use the following notation:

Definition 13.3 The function $H = H(\lambda, \beta)$ is called the *generalized inaccuracy function*, and $I = I(\lambda, \beta) := H(\lambda, \beta) - v^*(\lambda)$ is called the *generalized divergence function*.

13.1.2 Stability/Instability Properties

As shown by the following examples, there are families $(x_\beta^\varepsilon)_{\varepsilon>0}$ of ε-optimal decisions x_β^ε with respect to β, hence, $x_\beta^\varepsilon \in D_\varepsilon(\beta)$, $\varepsilon > 0$, such that

$$v(\lambda, x_\beta^\varepsilon) \geq H_0(\lambda, \beta) + \delta \quad \text{for all } 0 < \varepsilon < \varepsilon_0, \tag{13.15}$$

with a fixed constant $\delta > 0$ and for a positive ε_0.

Thus, with a certain distance $\delta > 0$, the expected loss remains—*also for arbitrarily small* accuracy value $\varepsilon > 0$ − outside the *error interval* $[v^*(\lambda), H_0(\lambda, \beta)]$, which must be taken into account in any case due to the estimation of (13.2a) by the approximate optimization problem (13.3). However, this indicates a possible *instability* of the 2-step procedure **(I, II)**.

Example 13.1 Let $\Omega := \{\omega_1, \omega_2\}$ with discrete probability distributions $\lambda, \beta \in \mathbb{R}^2_{+,1} := \{\lambda \in \mathbb{R}^2_+ : \lambda_1 + \lambda_2 = 1\}$. Moreover, define the set of decisions, the loss set, resp., by $D \equiv V$, where

$$V = conv\{(1,0)^T, (2,0)^T, (0,2)^T, (0,1)^T\} \backslash conv\{(\frac{1}{2}, \frac{1}{2})^T, (0,1)^T\},$$

where "conv" denotes the convex hull of a set. Selecting $\lambda = (0,1)^T$ and $\beta = (\frac{1}{2}, \frac{1}{2})^T$, we get

$$v^*(\beta) = \frac{1}{2}, H_\varepsilon(\lambda, \beta) = v(\lambda, x_\beta^\varepsilon) = 2(v^*(\beta) + \varepsilon) = 1 + 2\varepsilon$$

with $x_\beta^\varepsilon = (0, 2(v^*(\beta) + \varepsilon))^T = (0, 1 + 2\varepsilon)^T \in D_\varepsilon(\beta)$ for $0 < \varepsilon < \frac{1}{2}$. On the other hand, $D_0(\beta) = conv\{(\frac{1}{2}, \frac{1}{2})^T, (1,0)^T\} \backslash \{(\frac{1}{2}, \frac{1}{2})^T\}$, and therefore $H_0(\lambda, \beta) = \frac{1}{2}$. Hence, (13.15) holds, i.e., $H_\varepsilon(\lambda, \beta) = (\lambda, x_\beta^\varepsilon) = 1 + 2\varepsilon > 1 = H_0(\lambda, \beta) + \delta, \varepsilon > 0$, with $\delta = \frac{1}{2}$.

Remark 13.5 As it turns out later, the instability (13.15) follows from the fact that V is not closed.

Example 13.2 Let $\Omega = \{\omega_1, \omega_2\}$, and suppose that $D \equiv V$ is given by $V = \{(0,0)^T\} \cup \{z \in \mathbb{R}^2_+ : z_1 z_2 \geq 1\}$. Moreover, $\beta = (1,0)^T$ and $\lambda \in \mathbb{R}^2_{+,1}$ with $\lambda_2 > 0$. Then, $v^*(\beta) = 0$ and $D_0(\beta) = \{(0,0)^T\}$. Furthermore, $H_\varepsilon(\lambda, \beta) = +\infty$, and $x_\beta^\varepsilon = (\varepsilon, \frac{1}{\varepsilon})^T \in D_\varepsilon(\beta)$, where $v(\lambda, x_\beta^\varepsilon) = \lambda_1 \varepsilon + \lambda_2 \frac{1}{\varepsilon}$, $\varepsilon > 0$. Thus, $H_0(\lambda, \beta) = 0$, and also (13.15) holds

$$v(\lambda, x_\beta^\varepsilon) = \lambda_1 \varepsilon + \lambda_2 \frac{1}{\varepsilon} > H_0(\lambda, \beta) + \delta = \delta$$

for all $0 < \varepsilon < \frac{\lambda_2}{\delta}$ and each (fixed) $\delta > 0$.

Remark 13.6 Here, V is closed, but it is not convex, which is the reason for the instability (13.15) in the present case.

A necessary and sufficient condition excluding the *instability* (13.15) of the two-step procedure (I, II) procedure:

$$(13.1) \equiv (13.2a) \rightsquigarrow (13.1) \rightsquigarrow \text{select an} x_\beta^\varepsilon \in D_\varepsilon(\beta)$$

is given in the following result.

Lemma 13.2 *The* instability *(13.15) of the two-step procedure* (I, II) *is excluded if and only if* $H_\varepsilon(\lambda, \beta) \downarrow H_0(\lambda, \beta)$, $\varepsilon \downarrow 0$, *hence,* $H(\lambda, \beta) = H_0(\lambda, \beta)$.

Proof

(I) Suppose that $H(\lambda, \beta) = H_0(\lambda, \beta)$. Assuming that (13.15) holds, then $H_0(\lambda, \beta) < +\infty$ and therefore $H(\lambda, \beta) \in \mathbb{R}$ as well as $H_0(\lambda, \beta) < H_0(\lambda, \beta) + \delta \leq v(\lambda, x_\beta^\varepsilon) \leq H_\varepsilon(\lambda, \beta), \varepsilon > 0$. However, this is a contradiction to $H_\varepsilon(\lambda, \beta) \downarrow H_0(\lambda, \beta)$ for $\varepsilon \downarrow 0$. Consequently, (13.15) is excluded in this case.

(II) Suppose now that the instability (13.15) is excluded. Assuming that $H(\lambda, \beta) > H_0(\lambda, \beta)$, then $H_0(\lambda, \beta) \in \mathbb{R}$, and there is $c \in \mathbb{R}$, such that $H(\lambda, \beta) > c > H_0(\lambda, \beta)$. Because of $H_\varepsilon(\lambda, \beta) \geq H(\lambda, \beta) > c, \varepsilon > 0$, to each $\varepsilon > 0$ there exists an $x_\beta^\varepsilon \in D_\varepsilon(\beta)$ such that $v(\lambda, x_\beta^\varepsilon) > c$. Hence, (13.15) holds with $\delta := c - H_0(\lambda, \beta)$, which is in contradiction to the assumption. Consequently, we have $H(\lambda, \beta) = H_0(\lambda, \beta)$. □

In the following we give now sufficient conditions for the stability condition $H(\lambda, \beta) = H_0(\lambda, \beta)$ or $H_\varepsilon(\lambda, \beta) \downarrow H_0(\lambda, \beta)$ for $\varepsilon \downarrow 0$.

Theorem 13.1

(I) Let $D_\varepsilon(\beta) \neq \emptyset, \varepsilon > 0$, *and suppose that there is* $\bar{\varepsilon} > 0$ *such that* $D_{\bar{\varepsilon}}(\beta)$ *is compact und* $x \to v(\lambda, x), x \to v(\beta, x), x \in D_{\bar{\varepsilon}}(\beta)$ *are real valued, continuous functions on* $D_{\bar{\varepsilon}}(\beta)$. *Then,* $v^*(\beta) \in \mathbb{R}$, $D_0(\beta) \neq \emptyset$, *and* $H(\lambda, \beta) = H_0(\lambda, \beta)$.

(II) Replacing $x \to v(\lambda, x), x \in D_{\bar{\varepsilon}}(\beta)$, *by an arbitrary continuous function* $F : D_{\bar{\varepsilon}}(\beta) \to \mathbb{R}$ *and assuming that the remaining assumptions of the first part are unchanged, then* $\sup\{F(x) : x \in D_\varepsilon(\beta)\} \downarrow \sup\{F(x) : x \in D_0(\beta)\}$ *for* $\varepsilon \downarrow 0$.

Proof Obviously, with $F(x) := v(\lambda, x)$ the first assertion follows from the second one. Thus, we have to prove only the second part of Theorem 13.1. We therefore set $F_\varepsilon := sup\{F(x) : x \in D_\varepsilon(\beta)\}$ for $\varepsilon \geq 0$. Corresponding to Lemma 13.1, one can prove that $F_{\varepsilon_1} \leq F_{\varepsilon_2}$, provided that $\varepsilon_1 < \varepsilon_2$. In the first part of the proof we show that $D_0(\beta) \neq \emptyset$, hence, the expression $H_0(\lambda, \beta)$ is defined therefore. Since $D_{\bar{\varepsilon}}(\beta) \neq \emptyset$ and $v(\beta, x) \in \mathbb{R}$ for all $x \in D_{\bar{\varepsilon}}(\beta)$, we get $v^*(\beta) \geq -\bar{\varepsilon} + v(\beta, x) > -\infty$ with some $x \in D_{\bar{\varepsilon}}(\beta)$. Thus, $v^*(\beta) \in \mathbb{R}$. Assuming that $D_0(\beta) = \emptyset$, we would have $\cap_{0<\varepsilon\leq\bar{\varepsilon}} D_\varepsilon(\beta) = D_0(\beta) = \emptyset$. However, according to our assumptions, the set $D_\varepsilon(\beta) = \{x \in D_{\bar{\varepsilon}}(\beta) : v(\beta, x) \leq v^*(\beta) \leq v^*(\beta) + \varepsilon\}, 0 < \varepsilon \leq \bar{\varepsilon}$ is closed for each $0 < \varepsilon \leq \bar{\varepsilon}$. Due to the compactness of $D_{\bar{\varepsilon}}(\beta)$, this yields then $\cap_{i=1}^n D_{\varepsilon_i}(\beta) = \emptyset$ for a finite number of $0 < \varepsilon_i \leq \bar{\varepsilon}, i = 1, 2, ..., n$. Defining $\varepsilon_0 = \min\limits_{1\leq i\leq n} \varepsilon_i$, then $\varepsilon_0 > 0$ and $D_{\varepsilon_0}(\beta) = \cap_{i=1}^n D_{\varepsilon_i}(\beta) = \emptyset$, which contradicts to $D_\varepsilon \neq \emptyset, \varepsilon > 0$. Thus, we must have

$D_0(\beta) \neq \emptyset$. Since the sets $D_\varepsilon(\beta)$, $0 \leq \varepsilon \leq \bar{\varepsilon}$ are closed and therefore also compact, we have

$$F_\varepsilon = \sup\{F(x) : x \in D_\varepsilon(\beta)\} = \max\{F(x) : x \in D_\varepsilon(\beta)\}, \quad 0 \leq \varepsilon \leq \bar{\varepsilon}, \tag{13.16}$$

where, because of the continuity of F, the maximum is taken. In the second part we show that $F_\varepsilon \downarrow F_0$ for $\varepsilon \downarrow 0$. Due to the monotonicity of $\varepsilon \to F_\varepsilon$ and $F_\varepsilon \geq F_0$, $\varepsilon > 0$, we have $\lim\limits_{\varepsilon \downarrow 0} F_\varepsilon \geq F_0$. Moreover, with (13.16), for $0 \leq \varepsilon \leq \bar{\varepsilon}$ and each $c \in R$ it holds

$$\triangle_c := \{\varepsilon : 0 \leq \varepsilon \leq \bar{\varepsilon}, F_\varepsilon \geq c\} = \{\varepsilon : 0 \leq \varepsilon \leq \bar{\varepsilon}, \max\{F(x) : x \in D_\varepsilon(\beta)\} \geq c\} \tag{13.17a}$$

and therefore

$$\triangle_c = \{\varepsilon : 0 \leq \varepsilon \leq \bar{\varepsilon}, \text{ there is } x \in D_{\bar{\varepsilon}}(\beta) \text{ with } v(\beta, x) \leq v^*(\beta) + \varepsilon \text{ and } F(x) \geq c\}. \tag{13.17b}$$

If $\varepsilon^k \to \varepsilon^0, k \to \infty$ is a convergent sequence in $\triangle_c$ for a fixed $c \in \mathbb{R}$, then, according to (13.17a), (13.17b), there are elements $x^k \in D_{\bar{\varepsilon}}(\beta)$, such that

$$v(\beta, x^k) \leq v^*(\beta) + \varepsilon_k \text{ and } F(x^k) \geq c, \quad k = 1, 2, \ldots. \tag{13.18}$$

Since $D_{\bar{\varepsilon}}(\beta)$ is compact, sequence (x^k) has an accumulation point $x^0 \in D_{\bar{\varepsilon}}(\beta)$, and the continuity of $x \to F(x)$ and $x \to v(\beta, x)$ on $D_{\bar{\varepsilon}}(\beta)$ yield the existence of a subsequence (x^{k_j}) of (x^k), such that $F(X^{k_j}) \to F(x^0)$ and $v(\beta, x^{k_j}) \to v(\beta, x^0)$, $j \to \infty$. From (13.18) we get then $v(\beta, x^0) \leq v^*(\beta) + \varepsilon^0$ and $F(x^0) \geq c$. Hence, $F_{\varepsilon^0} \geq c$ and therefore $\varepsilon^0 \in \triangle_c$, because we have $0 \leq \varepsilon \leq \bar{\varepsilon}$, since $0 \leq \varepsilon^{k_j} \leq \bar{\varepsilon}$. The above considerations yield that $\triangle_c$ is closed for all $c \in R$. Assuming that there is $\tilde{c} \in R$, such that $\lim\limits_{\varepsilon \downarrow 0} F_\varepsilon \geq \tilde{c} > F_0$,, we have $F_\varepsilon \geq \tilde{c} > F_0$ for all $\varepsilon > 0$. However, this yields then $\triangle_{\tilde{c}} = \{0 \leq \varepsilon \leq \bar{\varepsilon} : F_\varepsilon \geq \tilde{c} :\} = (0, \bar{\varepsilon}]$, which contradicts to the closedness of $\triangle_c$ for each $c \in R$ shown above.

Thus, $\lim\limits_{\varepsilon \downarrow 0} F_\varepsilon = F_0$. □

Remark 13.7 According to (13.1)–(13.3), the second part of Theorem 13.1 is used mainly for $F(x) = -v(\lambda, x)$. In this case we have

$$\inf\{v(\lambda, x) : x \in D_\varepsilon(\beta)\} \uparrow \inf\{v(\lambda, x) : x \in D_0(\beta)\} \quad \text{for } \varepsilon \downarrow 0. \tag{13.19}$$

Obviously, the above result can be formulated also with the loss set V in the following way:

Corollary 13.1

(I) Let $V_\varepsilon(\beta) \neq \emptyset, \varepsilon > 0$ and suppose that there is $\bar{\varepsilon} > 0$ such that $V_{\bar{\varepsilon}}(\beta)$ is compact and $f \to \lambda f$, $f \to \beta f$, $f \in V_{\bar{\varepsilon}}(\beta)$ are real valued and continuous functions on $V_{\bar{\varepsilon}}(\beta)$. Then $v^*(\beta) \in R$, $V_0(\beta) \neq \emptyset$, and $H(\lambda, \beta) = H_0(\lambda, \beta)$;

(II) Replacing $f \to \lambda f$, $f \in V_{\bar{\varepsilon}}(\beta)$ *by an arbitrary continuous function* $F : V_{\bar{\varepsilon}}(\beta) \to R$, *and keeping the remaining assumptions in (I), then* $sup\{F(f) : f \in V_\varepsilon(\beta)\} \downarrow sup\{F(f) : f \in V_0(\beta)\}$ *for* $\varepsilon \downarrow 0$.

Proof The assertion follows immediately from $V_\varepsilon(\beta) = \{v(., x) : x \in D_\varepsilon(\beta)\}$. □

13.2 The Generalized Inaccuracy Function $H(\lambda, \beta)$

Let denote $P_{\tilde{\omega}} = \lambda$ the true distribution of $\tilde{\omega}$, and suppose that the hypothesis "$P_{\tilde{\omega}} = \beta$" has been accepted. Moreover, assume that $D_\varepsilon(\beta) \neq \emptyset$ for all $\varepsilon > 0$; This holds if and only if $v^*(\beta) > -\infty$ or $v^*(\beta) = -\infty$ and then $v(\beta, x) = -\infty$ for an $x \in D$. Using a decision $x \in D_\varepsilon(\beta)$, then we have a loss from $\{v(\lambda, x) : x \in D_\varepsilon(\beta)\}$, and

$$H_\varepsilon(\lambda, \beta) = \sup\{v(\lambda, x) : x \in D_\varepsilon(\beta)\},$$
$$h_\varepsilon(\lambda, \beta) = \inf\{v(\lambda, x) : x \in D_\varepsilon(\beta)\}$$

denotes the maximum, minimum, resp., expected loss, if the computation of an ε-optimal decision is based on the hypothesis "$P_{\tilde{\omega}} = \beta$", while $P_{\tilde{\omega}} = \lambda$ is the true probability distribution of $\tilde{\omega}$. Corresponding to Lemma 13.1 on $H_\varepsilon(\lambda, \beta)$, we can show this result:

Lemma 13.3 *Suppose that* $D_\varepsilon(\beta) \neq \emptyset$, $\varepsilon > 0$. *Then,*

(I) $\varepsilon \to h_\varepsilon(\lambda, \beta)$, *with* $h_\varepsilon(\lambda, \beta) = \inf\{v(\lambda, x) : x \in D_\varepsilon(\beta)\}$, *is monotonous decreasing on* $(0, +\infty)$;
(II) $h_\varepsilon(\lambda, \beta) \geq v^*(\lambda)$ *for all* $\varepsilon > 0$;
(III) $\varepsilon \to h_\varepsilon(\lambda, \beta)$, $\varepsilon > 0$ *is convex, provided that the loss set V is convex;*
(IV) If $v(\beta, x) \leq v^*(\beta) + \bar{\varepsilon}$ *for all* $x \in D$ *and a fixed* $\bar{\varepsilon} > 0$, *then* $h_\varepsilon(\lambda, \beta) = v^*(\lambda)$, $\varepsilon > \bar{\varepsilon}$;
(V) The assertions (a)–(c) hold also for $\varepsilon \geq 0$, *in case that* $D_o(\beta) \neq \emptyset$.

Lemmas 13.1 and 13.3 yield then this corollary:

Corollary 13.2 *For two numbers* $\varepsilon_1, \varepsilon_2 > 0$ (≥ 0, *if* $D_o(\beta) \neq \emptyset$, *resp.) we have*

$$h_{\varepsilon_1}(\lambda, \beta) \leq H_{\varepsilon_2}(\lambda, \beta). \tag{13.20}$$

Proof If $\varepsilon_1 \leq \varepsilon_2$, then $H_{\varepsilon_2}(\lambda, \beta) \geq H_{\varepsilon_1}(\lambda, \beta) \geq h_{\varepsilon_1}(\lambda, \beta)$, and in case $\varepsilon_1 > \varepsilon_2$ it is $h_{\varepsilon_1}(\lambda, \beta) \leq h_{\varepsilon_2}(\lambda, \beta) \leq H_{\varepsilon_2}(\lambda, \beta)$ according to (a) of Lemmas 13.1, 13.3. □

Hence, the limit $\lim_{\varepsilon \downarrow 0} h_\varepsilon(\lambda, \beta) = \sup_{\varepsilon > 0} h_\varepsilon(\lambda, \beta)$, exists, and corresponding to (13.14a) we define

$$h(\lambda, \beta) = \lim_{\varepsilon \downarrow 0} h_\varepsilon(\lambda, \beta) = \sup_{\varepsilon > 0} h_\varepsilon(\lambda, \beta). \tag{13.21}$$

For the functions $h(\lambda, \beta)$, $H(\lambda, \beta)$ defined by (13.14a) and (13.21) the following result holds:

Theorem 13.2

(I) Let $D_\varepsilon(\beta) \neq \emptyset$ for $\varepsilon > 0$. Then

$$v^*(\lambda) \leq h(\lambda, \beta) \leq H(\lambda, \beta);$$

(II) Let $D_\varepsilon(\lambda) \neq \emptyset$ for $\varepsilon > 0$. Then

$$v^*(\lambda) = h(\lambda, \lambda) = H(\lambda, \lambda).$$

Proof

(I) Lemma 13.3 yields $h(\lambda, \beta) = \sup_{\varepsilon>0} h_\varepsilon(\lambda, \beta) \geq v^*(\lambda)$. Because of (13.20) we have $h_\varepsilon(\lambda, \beta) \leq H_{\bar{\varepsilon}}(\lambda, \beta)$, $\varepsilon > 0$ for each fixed $\bar{\varepsilon} > 0$. From this we obtain $h(\lambda, \beta) = \sup_{\varepsilon>0} h_\varepsilon(\lambda, \beta) \leq H_{\bar{\varepsilon}}(\lambda, \beta)$, $\bar{\varepsilon} > 0$, hence, $h(\lambda, \beta) \leq \inf_{\varepsilon>0} H_\varepsilon(\lambda, \beta) = H(\lambda, \beta)$.

(II) According to Theorem 13.2 (I) and the Definition of $H(\lambda, \lambda)$, we get $v^*(\lambda) \leq H(\lambda, \lambda) \leq H_\varepsilon(\lambda, \lambda) = \sup\{v(\lambda, x) : x \in D_\varepsilon(\lambda)\} = \sup\{v(\lambda, x) : v(\lambda, x) \leq v^*(\lambda) + \varepsilon\} \leq v^*(\lambda) + \varepsilon$ for each $\varepsilon > 0$. Thus, $H(\lambda, \lambda) = v^*(\lambda)$, and due to the first part we also have $v^*(\lambda) = h(\lambda, \lambda)$. □

For the geometrical interpretation of the values $h(\lambda, \beta)$, $H(\lambda, \beta)$ consider now the transformed loss set

$$V_{\lambda,\beta} = \{(\lambda f, \beta f)^T : f \in V\} (\subset \mathbb{R}^2) . \tag{13.22}$$

If the loss set V is convex, then the linear transformation $V_{\lambda,\beta} = T_{\lambda,\beta}(V)$ of V with respect to $T_{\lambda,\beta} : f \to (\lambda f, \beta f)^T$ is again a convex set. Hence, $v^0(\beta) = \inf\{z_2 : z \in V_{\lambda,\beta}\}$, which means that $v^*(\beta)$ can be interpreted as the second coordinate of one of the deepest points of $V_{\lambda,\beta}$. In the same way, $v^*(\lambda)$ is the first coordinate one of the points lying on the left boundary of $V_{\lambda,\beta}$.

According to Fig. 13.2, the values $h(\lambda, \beta)$, $H(\lambda, \beta)$ and also the divergences $I(\lambda, \beta) := H(\lambda, \beta) - H(\lambda, \lambda)$, $J(\lambda, \beta) := h(\lambda, \beta) - h(\lambda, \lambda)$ can be interpreted corresponding to $V_{\lambda,\beta}$ in this way:

$$\begin{aligned} h(\lambda, \beta) &:= \text{ first coordinate of the deepest point of } V_{\lambda,\beta} \text{ being most left} \\ H(\lambda, \beta) &:= \text{ first coordinate of the deepest point of } V_{\lambda,\beta} \text{ being most right.} \end{aligned}$$

The remaining H, h- and I, J-functions can be interpreted in $V_{\lambda,\beta}$ in the same way.

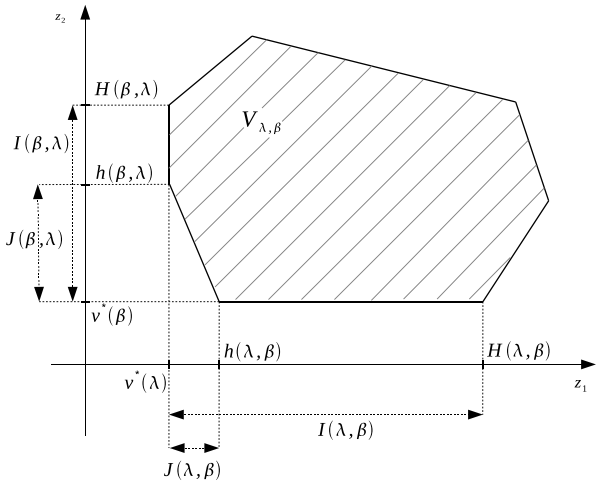


Fig. 13.2 The H, h- and I, J-functions

13.2.1 Special Loss Sets V

In the following we give a justification for the notation "*generalized inaccuracy function*" for $H(\lambda, \beta)$ and $h(\lambda, \beta)$. For this aim, assume next to that $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ contains a finite number of realizations or scenarios. Moreover, suppose that $f : \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$ is a convex, monotonous decreasing function such that $f(t) \in \mathbb{R}$ for $t > 0$ and $f(0) = \lim_{t\to 0} f(t) = \sup_{t>0} f(t)$. Putting $f(\alpha) \equiv (f(\alpha_1), f(\alpha_2), ..., f(\alpha_n))^T$ and $ri\mathbb{R}^n_{+,1} = \{\alpha \in \mathbb{R}^n_{+,1} : \alpha_k > 0, k = 1, 2, ..., n\}$, let then the loss set

$$V := C_f,$$

be defined by

$$C_f = clconv\{f(\alpha) : \alpha \in ri\mathbb{R}^n_{+,1}\};. \tag{13.23}$$

Here, "$clconv$" denotes the closed, convex hull of a set. We still put

$$v^*_f(\beta) = \inf\{\beta^T z : z \in C_f\}, \beta \in \mathbb{R}^n_{+,1}.$$

Some properties of C_f are stated in the following:

Lemma 13.4

(I) *C_f is closed and convex;*
(II) *From $z \in C_f$ we also have $(z_{\tau(1)}, ..., z_{\tau(n)})^T \in C_f$ for each permutation τ of the index set $\{1, 2, ..., n\}$;*

(III) *If* $f(t) \geq 0, 0 \leq t \leq 1$, *then* $C_f \subset \mathbb{R}^n_+, 0 \leq v^*_f(\beta) < +\infty$, *and for* $\beta_k > 0$ *we get* $0 \leq z_k \leq (\frac{1/\beta}{k})(v^*_f(\beta) + \varepsilon), z \in V_\varepsilon(\beta)$;

(IV) *If* $f(0) \in \mathbb{R}$, *then* C_f *is a compact, convex subset of* $\mathbb{R}^n$, *it also holds* $C_f = conv\{f(\alpha) : \alpha \in \mathbb{R}^n_{+,1}\}$, *and to each* $z \in C_f$ *there is an* $\alpha \in \mathbb{R}^n_{+,1}$ *with* $z \geq f(\alpha)$;

(V) *If* $f(0) = +\infty$, *then for each* $z \in C_f$ *there exists an* $\alpha \in ri\mathbb{R}^n_{+,1}$ *such that* $z \geq f(\alpha)$.

Proof

(I) The first part follows from the definition of C_f.

(II) Each $z \in C_f$ has the representation $z = \lim_{\nu\to\infty} z^\nu$ with $z^\nu = \sum_{i=1}^{n} \nu_\gamma^{\nu i} \cdot f(\alpha^{\nu i}), \alpha^{\nu i} \in riR^n_{+,1}, \gamma^{\nu i} \geq 0, i = 1, 2, ..., n_\nu, \sum_{i=1}^{n} \nu_\gamma^{\nu i} = 1$. Consequently, with a permutation τ of $1, 2, ..., n$, also $(x_{\tau(1)}, ..., x_{\tau(n)})^T$ has a representation of this type. Thus, $(x_{\tau(1)}, ..., x_{\tau(n)})^T \in C_f$.

(III) From $f(t) \geq, 0 \leq t \leq 1$ and $0 < \alpha_k < 1$ for $\alpha \in ri\mathbb{R}^n_{+,1}$, we get $f(\alpha) \in \mathbb{R}^n_+$ for $\alpha \in ri\mathbb{R}^n_{+,1}$ and therefore $C_f \subset \mathbb{R}^n_+$. Hence, $v^*_f(\beta) = \inf\{\beta^T z : z \in C_f\} \geq 0$ and $z \geq 0$. Because of $(f(1/n), ..., f(1/n))^T \in C_f$ and $f(1/n) \in \mathbb{R}$, we find $v^*_f(\beta) \leq \sum_{k=1}^{n} \beta_k f(1/n) < +\infty$. In addition, because of $z_k \geq 0, k = 1, 2, ..., n$ for $z \in C_f$ and with $\beta \geq 0$ we get $z_k\beta_k \leq \beta^T z \leq v^*(\beta) + \varepsilon$ for each $z \in V_\varepsilon(\beta)$, hence, $0 \leq z_k \leq (1/\beta_k)(v^*_f(\beta) + \varepsilon)$, provided that $\beta_k > 0$.

(IV) Since $\{f(\alpha) : \alpha \in ri\mathbb{R}^n_{+,1}\} \subset \{f(\alpha) : \alpha \in \mathbb{R}^n_{+,1}\}$ and $\alpha \to f(\alpha) = (f(\alpha_1), ..., f(\alpha_n))^T, \alpha \geq 0$ is a continuous mapping for real $f(0)$, we find that $\{f(\alpha) : \alpha \in ri\mathbb{R}^n_{+,1}\}$ is bounded as a subset of the compact set $\{f(\alpha) : \alpha \in ri\mathbb{R}^n_{+,1}\}$. Due to [14], Theorem 17.2 we obtain then $C_f = clconv\{f(\alpha) : \alpha \in ri\mathbb{R}^n_{+,1}\} = conv(cl\{f(\alpha) : \alpha \in ri\mathbb{R}^n_{+,1}\}) = conv\{f(\alpha); \alpha \in \mathbb{R}^n_{+,1}\}$; indeed, if $f(\alpha^\nu) \to z, \nu \to \infty$ with $\alpha^\nu \in ri\mathbb{R}^n_{+,1}$, then, due to the compactness of $\mathbb{R}^n_{+,1}$ we have a subsequence (α^{ν_j}) of (α^ν), such that $\alpha^{\nu_j} \to \alpha \in \mathbb{R}^n_{+,1}, j \to \infty$. Because of $f(\alpha^{\nu_j}) \to z, j \to \infty$ and the continuity of f, we get then $f(\alpha) = z$, hence, $z \in \{f(\alpha) : \alpha \in \mathbb{R}^n_{+,1}\}$ and therefore, as asserted, $cl\{f(\alpha) : \alpha \in ri\mathbb{R}^n_{+,1}\} = \{f(\alpha) : \alpha \in \mathbb{R}^n_{+,1}\}$. Being the convex hull of a compact set, C_f is also a compact set. For $z \in C_f$ we have $z = \sum_{i=1}^{\nu} \gamma^i f(\alpha^i), \gamma^i \geq 0, \alpha^i \in \mathbb{R}^n_{+,1}, i = 1, 2, ...,$ and $\sum_{i=1}^{\nu} \gamma^i = 1$. Thus, $z_k = \sum_{i=1}^{\nu} \gamma^i f(\alpha^i_k) \geq f(\sum_{i=1}^{\nu} \gamma^i \alpha^i_k) = f(\alpha_k)$ with $\alpha = \sum_{i=1}^{\nu} \gamma^i \alpha^i \in \mathbb{R}^n_{+,1}$. Hence, we have therefore found an $\alpha \in \mathbb{R}^n_{+,1}$ such that $z \geq f(\alpha)$.

(V) Due to the representation of an element $z \in C_f$ stated in part (II), we have $z_k = \lim_{\nu\to\infty} z^\nu_k$, where $z^\nu_k = \sum_{i=1}^{n_\nu} \gamma^{\nu i} f(\alpha^{\nu i}_k)$ with $\alpha^{\nu i} \in ri\mathbb{R}^n_{+,1}$ and $\gamma^{\nu i} \geq 0, \sum_{i=1}^{n_\nu} \gamma^{\nu i} = 1$. As above, for each $\nu = 1, 2, ...$ we have the relation $z^\nu \geq f(\alpha^\nu)$, with $\alpha^\nu = \sum_{i=1}^{n} \gamma^{\nu i} \alpha^{\nu i}$. However, the sequence (α^ν) has an accumulation point α in

$\mathbb{R}^n_{+,1}$; We show now that $\alpha \in ri\mathbb{R}^n_{+,1}$. Assuming that $\alpha_k = 0$ for an index $1 \leq k \leq n$, with a sequence $\alpha^{\nu_j} \to \alpha$, $j \to \infty$ we get the relation $\alpha_k^{\nu_j} \to \alpha_k = 0$ and therefore $f(\alpha_k^{\nu_j}) \to$ f(0)$= +\infty$, $j \to \infty$. However, this is not possible, since $z_k^{\nu_j} \geq f(\alpha_k^{\nu_j})$ and (z^ν) is a convergent sequence. Thus, $\alpha_k > 0$. Furthermore, from $z_k^{\nu_j} \geq f(\alpha_k^{\nu_j})$, $j = 1, 2, ..., z_k^{\nu_j} \to z_k, \alpha_k^{\nu_j} \to \alpha_k$, $j \to \infty$ we finally obtain $z_k \geq f(\alpha_k)$, hence, $z \geq f(\alpha)$ with an $\alpha \in ri\mathbb{R}^n_{+,1}$. □

The above lemma yields now several consequences on $H_\varepsilon(\lambda, \beta)$, $h_\varepsilon(\lambda, \beta)$.

Corollary 13.3 *For each $\lambda \in \mathbb{R}^n_{+,1}$ the value $v_f^*(\lambda)$ has the representation*

$$v_f^*(\lambda) = \inf\{\lambda^T f(\alpha) : \alpha \in ri R^n_{+,1}\} \tag{13.24}$$

and for $f(0) \in \mathbb{R}$ we may replace $ri\mathbb{R}^n_{+,1}$ also by $\mathbb{R}^n_{+,1}$.

For $H(\lambda, \beta) = H^{(f)}(\lambda, \beta)$ and $h(\lambda, \beta) = h^{(f)}(\lambda, \beta)$, with $V = C_f$, from Lemma 13.4 we get this result:

Corollary 13.4

(I) If $f(0) \in \mathbb{R}$, then $H^{(f)}(\lambda, \beta) = H_0^{(f)}(\lambda, \beta)$ and $h^{(f)}(\lambda, \beta) = h_0^{(f)}(\lambda, \beta) = \inf\{\lambda^T f(\alpha) : \beta^T f(\alpha) = v_f^*(\beta)\}$ for all $\lambda, \beta \in \mathbb{R}^n_{+,1}$. Furthermore, $H_0^{(f)}(\lambda, \beta) = \sup\{\lambda^T f(\alpha) : \beta^T f(\alpha) = v_f^*(\beta)\}$ for all $\lambda \in \mathbb{R}^n_{+,1}$ and $\beta \in ri\mathbb{R}^n_{+,1}$.

(II) If $f(0) = +\infty$ and $f(t) \geq 0$, $0 \leq t \leq 1$, then $H^{(f)}(\lambda, \beta) = H_0^{(f)}(\lambda, \beta)$, $h^{(f)}(\lambda, \beta) = h_0^{(f)}(\lambda, \beta)$, provided that $\lambda \in \mathbb{R}^n_{+,1}$ and $\beta \in ri\mathbb{R}^n_{+,1}$. If $\lambda, \beta \in \mathbb{R}^n_{+,1}$ are selected such that $\lambda_k > 0$, $\beta_k = 0$ for an $1 \leq k \leq n$, then $H(\lambda, \beta) = +\infty$.

For the case $\beta \in \mathbb{R}^n_{+,1} \backslash ri\mathbb{R}^n_{+,1}$ we obtain this result:

Corollary 13.5 *If $\beta_\kappa = 0$ for an index $l \leq \kappa \leq n$, then*

$$v_f^*(\beta) = \inf\{\sum_{k=\kappa}^n \beta_k f(\alpha_k) : \alpha_k > 0, k \neq \kappa, \sum_{k \neq \kappa}^n \alpha_k = 1\}, \tag{13.25}$$

and in the case $f(0) \in \mathbb{R}$ the inequality "$\alpha_k > 0$" may be replaced also by "$\alpha_k \geq 0$". If $f(0) = +\infty$ and in addition f is strictly monotonous decreasing on $[0, 1]$, then $V_0(\beta) = \emptyset$.

Indicating the dependence of the set C_f, cf. (13.23), as well as the values $v_f^*(\beta)$ on the index n (= number of elements of Ω) by means of $C_f^{(n)}$, $v_f^{(n)*}(\beta)$, resp., then $v_f^{(n)*}(\beta) = \inf\{\beta^T z : z \in C_f^{(n)}\}$. Moreover, for a $\beta \in \mathbb{R}^n_{+,1}$ with $\beta_k = 0$, $1 \leq k \leq n$, and using the notation $\hat{\beta} = (\beta_1, ..., \beta_{k-1}, \beta_{k+1}, ..., \beta_n)^T$, the equations (13.24) and (13.25) yield

$$v_f^{(n)*}(\beta) = v_f^{(n-1)*}(\hat{\beta}). \tag{13.26}$$

Extending (13.26), we find the following corollary:

Corollary 13.6 *Suppose that* $\lambda, \beta \in \mathbb{R}^n_{+,1}$ *with* $\lambda_k = \beta_k = 0$ *for an index* $1 \le k \le n$. *Moreover, the notation* $h^{(n)}_\varepsilon(\lambda, \beta), h^{(n)}(\lambda, \beta)$ *indicates the dependence of the functions* $h_\varepsilon(\lambda, \beta), h(\lambda, \beta)$ *on the dimension* n. *Then,* $h^{(n)}_\varepsilon(\lambda, \beta) = h^{(n-1)}_\varepsilon(\hat{\lambda}, \hat{\beta})$ *for* $\varepsilon > 0$, *as well as* $h^{(n)}(\lambda, \beta) = h^{(n-1)}(\hat{\lambda}, \hat{\beta})$, *provided that* $\hat{\lambda} = (\lambda_1, ..., \lambda_{k-1}, \lambda_{k+1}, ..., \lambda_n)'$, *and* $\hat{\beta}$ *is defined in the same way.*

A corresponding result for $H(\lambda, \beta)$ is stated below:

Corollary 13.7 *Consider* $\lambda, \beta \in \mathbb{R}^n_{+,1}$ *with* $\lambda_k = \beta_k = 0$ *for a certain* $l \le k \le n$, *and let indicate* $H^{(n)}(\lambda, \beta)$ *the dependence of* $H(\lambda, \beta)$ *on* n. *Then,* $H^{(n)}(\lambda, \beta) = H^{(n-1)}(\lambda, \beta)$, *provided that* λ, β *are defined as above, and the implication* $\beta_j = 0 \Rightarrow \lambda_j = 0$ *holds for* $j \ne k$.

A relationship between $v^*_f(\lambda)$ and $v^*_f(e), e = (1/n, ..., 1/n)^T$ is shown in the following.

Corollary 13.8 *For all* $\lambda \in R^n_{+,1}$ *we have* $v^*_f(\lambda) \le f(1/n) \le v^*_f(1/n, ..., 1/n)$.

Proof Equation (13.24) in Corollary 13.3 yields $v^*_f(\lambda) \le \lambda^T f(e) = f(1/n)$. Select then an arbitrary $\varepsilon > 0$. According to (13.24) there exists an $\alpha^\varepsilon \in ri\,\mathbb{R}^n_{+,1}$, such that $v^*(e) + \varepsilon \ge e^T f(\alpha^\varepsilon)$. Hence, $v^*(e) \ge -\varepsilon + \sum_{k=1}^n f(\alpha^\varepsilon_k) \ge -\varepsilon + f(\sum_{k=1}^n (1/n)\alpha^\varepsilon_k) = -\varepsilon + f(1/n)$. Since $\varepsilon > 0$ was chosen arbitrarily, we have $v^*_f(e) \ge f(1/n)$ and therefore $v^*_f(e) \ge f(1/n) \ge v^*_f(\lambda)$. □

Note that the assertion in this corollary can also be found in [1]. Having some properties of $H^{(n)}$ and $h^{(n)}$, we determine now these functions for some important

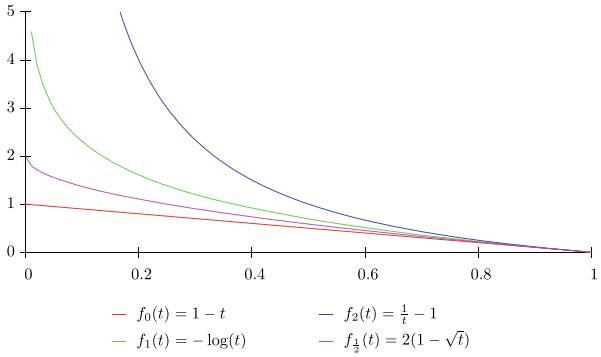


Fig. 13.3 Functions f_b

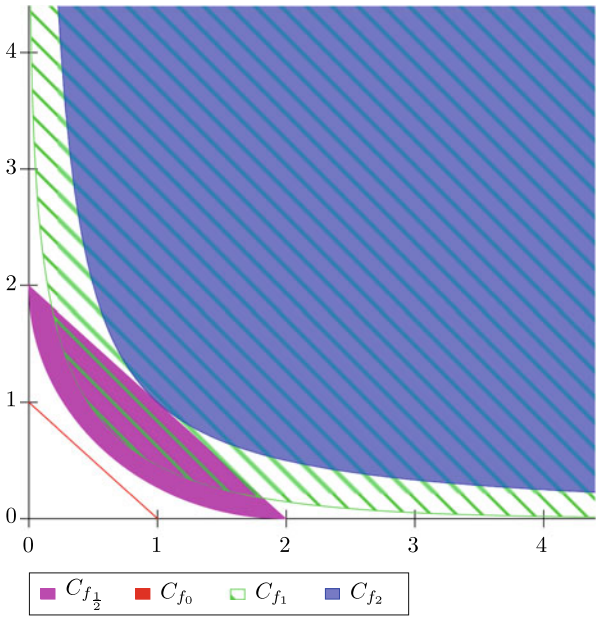


Fig. 13.4 Convex sets C_{f_b}

special cases of f. Next to we consider, see Fig. 13.3, the family $(f_b)_{b\geq 0}$, defined by (cf. [1])

$$f_b(t) = \begin{cases} \frac{1}{1-b}(1-t^{1-b}) \ , t \geq 0 & \text{for } b > 0, b \neq 1 \\ -\log t \qquad\quad , t \geq 0 & \text{for } b = 1. \end{cases}$$

The corresponding sets C_{f_b} are shown in the next Fig. 13.4.

It is easy to see that each f_b is a strictly monotonous decreasing, convex and for $b > 0$ strictly convex function on $[0, +\infty]$, such that $-\infty < f_b(t) < +\infty, t > 0$ and

$$\lim_{t\to 0} f_b(t) = \sup_{t>0} f_b(t) = f_b(0) = \begin{cases} \frac{1}{1-b} & \text{for } 0 \leq b \leq 1 \\ +\infty & \text{for } b \geq 1. \end{cases}$$

Moreover, $f_b(t) \geq 0 = f(1)$ for $0 \leq t \leq 1$. Hence, $f = f_b$ fulfills all needed conditions. Next to we want to determine $v_f^*(\lambda)$ and $V_o(\lambda)$, where the dependence on $b \geq$ is denoted by the notations $v_{(b)}^*(\lambda)$, $V_{(b)o}(\lambda)$ and $C_{(b)}$.

Theorem 13.3

(I) For each $\lambda \in \mathbb{R}^n_{+,1}$ we have

$$v^*_{(o)}(\lambda) = 1 - \max_{1\le k\le n} \lambda_k \tag{13.27a}$$

$$v^*_{(b)}(\lambda) = \frac{1}{1-b}\sum_{k=1}^{n} \lambda_k \left(1 - \left(\frac{\lambda_k 1/b}{\sum_{k=1}^{n} \lambda_k{}^{1/b}}\right)^{1-b}\right) \quad \text{for } b > 0, b \neq 1 \tag{13.27b}$$

$$v^*_{(1)}(\lambda) = \sum_{k=1}^{n} \lambda_k \log(1/\lambda_k) \tag{13.27c}$$

(II) If $\lambda_1 > 0, ..., \lambda_m > 0, \lambda_{m+1} = ... = \lambda_n = 0$, then for $b > 0$ we also have

$$v^*_{(b)}(\lambda) = \hat{\lambda}^T f_b(\hat{\alpha}^{(b)}) \tag{13.27d}$$

with $\lambda = (\lambda_1, ..., \lambda_m)^T$, $\hat{\alpha}^{(b)} = (\alpha_1^{(b)}, ..., \alpha_m^{(b)})^T$ and

$$\alpha_k^{(b)} = \begin{cases} \lambda_k^{1/b} / \sum_{k=1}^{n} \lambda_k^{1/b} & \text{for } b > 0, b \neq 1 \\ \lambda_k & \text{for } b = 1 \end{cases} \qquad \text{and } k = 1, ..., m,$$

where $\hat{\alpha}^{(b)}$ is determined uniquely.

(III) We have

$$V_{(o)o}(\lambda) = \{f_o(\alpha) : \alpha \in R^n_{+,1}, \sum_{k=1}^{n} \lambda_k \alpha_k = \max_{1\le k\le n} \lambda_k\}, \quad \lambda \in \mathbb{R}^n_{+,1}, \tag{13.27e}$$

$$V_{(b)o}(\lambda) = \{f_b(\alpha^{(b)})\} \quad \text{with } \alpha^{(b)} = (\alpha_1^{(b)})^T, \lambda \in ri\,\mathbb{R}^n_{+,1}, b > 0 \tag{13.27f}$$

$$\begin{aligned} V_{(b)o}(\lambda) &= \{z : z_k = f_b(\alpha_k^{(b)}), \lambda_k > 0 \text{ and } z_k = \frac{1}{1-b}, \lambda_k = 0\} \\ &= \{f_b(\tilde{\alpha}^{(b)})\}, \lambda \in \mathbb{R}^n_{+,1} \backslash ri\,\mathbb{R}^n_{+,1}, 0 < b < 1, \text{and certain } \tilde{\alpha}^{(b)} \end{aligned} \tag{13.27g}$$

$$V_{(b)o}(\lambda) = \emptyset, \quad \lambda \in R^n_{+,1} \backslash ri\, R^n_{+,1} \text{ and } b \geq 1. \tag{13.27h}$$

Proof Consider first the case $b = 0$. From (13.23) we easily find that $C_{(o)} = clconv\{f_0(\alpha) : \alpha \in ri\,\mathbb{R}^n_{+,1}\} = \{(1-\alpha_k)_{k=1,...,n} : \alpha \in \mathbb{R}^n_{+,1}\}$. Because of (13.24) we further have $v^*_{(o)}(\lambda) = \inf\{1 - \lambda^T\alpha : \alpha \in \mathbb{R}^n_{+,1}\} = 1 - \sup\{\lambda^T\alpha : \alpha \in \mathbb{R}^n_{+,1}\} = 1 - \max_{1\le k\le n} \lambda_k$. Hence, $V_{(o)o}(\lambda) = \{z \in C_{(o)} : \lambda^T\alpha = \max_{l\le k\le n} \lambda_k, \alpha \in \mathbb{R}^n_{+,1}\}$, which shows the

assertion for $b = 0$. Thus, let now $b > 0$ and $\lambda_k > 0, k = 1, ..., m, \lambda_{m+1} = ... = \lambda_n = 0$. From (13.24) and (13.26) we get then $v_{(b)}^{(n)*}(\lambda) = v_{(b)}^{(m)*}(\hat{\lambda}) = \inf\{\hat{\lambda}^T f_b(\hat{\alpha}) : \hat{\alpha} \in ri\mathbb{R}_{+,1}^m\}$ with $\hat{\lambda} = (\lambda_1, ..., \lambda_m)^T$ and $\hat{\alpha} = (\alpha_1, ..., \alpha_m)^T$. The Lagrangian of the convex optimization problem

$$\min \quad \hat{\lambda}^T f_b(\hat{\alpha}) \tag{13.28a}$$

$$\text{s.t.} \quad \hat{\alpha} \in ri\mathbb{R}_{+,1}^m \tag{13.28b}$$

is then given by $L(\hat{\alpha}, u) = \hat{\lambda}^T f_b(\hat{\alpha}) + u(\sum_{k=1}^m \alpha_k - 1)$. Moreover, the optimality conditions—without considering the constraints $\alpha_k > 0, k = 1, ..., m$—read

$$0 = \frac{\partial L}{\partial u} = \sum_{k=1}^m \alpha_k - 1 \tag{13.29a}$$

$$0 = \frac{\partial L}{\partial \alpha_k} = \lambda_k Df_b(\alpha_k) + u \tag{13.29b}$$

Inserting $Df_b(t) = -t^{-b}, t > 0$ for $b > 0, b \neq 1$ and $Df_1(t) = -1/t, t > 0$ for $b = 1$ into (13.29a), yields

$$\alpha_k = \alpha_k^{(b)} = \begin{cases} \lambda_k^{1/b} / \sum_{k=1}^m \lambda_k^{1/b} & \text{for } b > 0, b \neq 1 \\ \lambda_k & \text{for } b = 1 \end{cases} \qquad \text{and } k = 1, \ldots, m.$$

Since $\alpha_k^{(b)} > 0, k = 1, ..., m$, also the conditions $\alpha_k > 0, k = 1, ..., m$, hold. Moreover, $\hat{\alpha}^{(b)} = (\alpha_1^{(b)}, ..., \alpha_m^{(b)})^T$ is the unique solution of (13.28a), (13.28b), since $\alpha \to \lambda^T f(\alpha)$ is strictly convex on $ri\mathbb{R}_{+,1}^m$. Hence, $v_{(b)}^{(n)*}(\lambda) = v_{(b)}^{(m)*}(\lambda) = \hat{\lambda}^T f_b(\hat{\alpha}^{(b)})$. Using the convention $0 \cdot (+\infty) = 0$, yields the rest of (13.27b). In addition, this proves also part (II). For showing (III) and therefore also (I), let $\lambda \in ri\mathbb{R}_{+,1}^n$ and $z \in V_{(b)o}(\lambda), b > 0$. We remember that $f_b(0) = +\infty$ for $b \geq 1$ and $f_b(0) \in R$ for $0 \leq b < 1$. Part (II) yields then $\alpha^{(b)} = (\alpha_1^{(b)}, ..., \alpha_n^{(b)})^T \in riR_{+,1}^n$ and $v_{(b)}^{(n)*}(\lambda) = \lambda^T f_b(\alpha^{(b)})$, hence, $f_b(\alpha^{(b)}) \in V_{(b)o}(\lambda) (\neq \emptyset)$.

According to Lemma 13.4, for $0 < b < 1$, $b \geq 1$, resp., there is $\alpha \in \mathbb{R}_{+,1}^n$, $\alpha \in ri\mathbb{R}_{+,1}^n$,, resp., such that $z \geq f_b(\alpha)$ (to each $z \in V_{(b)o}(\lambda)$). This immediately yields $v_{(b)}^{(n)*}(\lambda) = \lambda^T z = \lambda^T f_b(\alpha)$, since $f_b(\alpha) \in C_{(b)}$. Thus, $z = f_b(\alpha)$, because of $\lambda_k > 0, k = 1, ..., n$. Assuming $\alpha \neq \alpha^{(b)}$, for $\tilde{\alpha} = \frac{1}{2}\alpha + \frac{1}{2}\alpha^{(b)}$ we get on the one hand $\tilde{\alpha} \in ri\mathbb{R}_{+,1}^n$, hence, $f_b(\tilde{\alpha}) \in C_{(b)}, b > 0$, and on the other hand we have $v_{(b)}^{(n)*}(\lambda) \leq \lambda^T f_b(\tilde{\alpha}) < \frac{1}{2}\lambda^T f_b(\alpha) + \frac{1}{2}\lambda^T f_b(\alpha^{(b)}) = v_{(b)}^{(n)*}(\lambda)$, since f_b is strictly convex for $b > 0$. Consequently, $\alpha = \alpha^{(b)}$ and therefore $V_{(b)o}(\lambda) = \{f_b(\alpha^{(b)})\}, b > 0, \lambda \in ri\mathbb{R}_{+,1}^n$. Now consider $\lambda_1 > 0, ..., \lambda_m > 0, \lambda_{m+1} = ... = \lambda_n = 0$. Again from part (II) we get $v_{(b)}^{(n)*}(\lambda) = v_{(b)}^{(m)*}(\hat{\lambda}) = \lambda^T f_b(\hat{\alpha}^{(b)})$. We put $\tilde{\alpha}^{(b)} = (\alpha_1^{(b)}, ..., \alpha_m^{(b)}, 0, ..., 0)^T$. Let $0 < b < 1$. Because of $f(0) = 1/(1-b) \in \mathbb{R}$, due to Lemma 13.4 we obtain

$C_{(b)} = conv\{f_b(\alpha) : \alpha \in \mathbb{R}^n_{+,1}\}$, hence, $f_b(\tilde{\alpha}^{(b)}) \in V_{(b)o}(\lambda)$, since $\tilde{\alpha}^{(b)} \in \mathbb{R}^n_{+,1}$ and $\lambda^T f_b(\tilde{\alpha}^{(b)}) = \hat{\lambda}^T f_b(\hat{\alpha}^{(b)}) = v^{(n)*}_{(b)}(\lambda)$.

Consider now $z \in V_{(b)o}(\lambda)$. According to Lemma 13.4, part V, there is $\alpha \in \mathbb{R}^n_{+,1}$ with $z \geq f_b(\alpha)$. Because of $\hat{\lambda}^T \hat{z} = \lambda^T z = v^{(n)*}_{(b)}(\lambda) = \lambda^T f_b(\alpha) = \hat{\lambda}^T f_b(\hat{\alpha})$ we have $f_b(\alpha) \in V_{(b)o}(\lambda)$ and $\hat{z} = f_b(\hat{\alpha})$, since $\lambda_k > 0, k = 1, 2, ..., m$. Assuming $\hat{\alpha} \neq \hat{\alpha}^{(b)}$ and considering then under this assumption $\gamma = \frac{1}{2}\alpha + \frac{1}{2}\tilde{\alpha}^{(b)}$, we get $f_b(\gamma) \in C_{(b)}$, since $\gamma \in \mathbb{R}^n_{+,1}$ and $v^{(n)*}_{(b)}(\lambda) \leq \lambda^T f_b(\gamma) = \hat{\lambda}^T f_b(\hat{\gamma}) < \frac{1}{2}\hat{\lambda}^T f_b(\hat{\alpha}) + \frac{1}{2}\hat{\lambda}^T f_b(\hat{\alpha}^{(b)}) = v^{(n)*}_{(b)}(\lambda)$. However, this yields a contradiction, hence, it holds $\hat{\alpha}^{(b)} = \hat{\alpha}$. Obviously, each $\alpha \in \mathbb{R}^n_{+,1}$ fulfilling this equation is contained in $V_{(b)o}(\lambda)$.

Thus, $V_{(b)o}(\lambda) = \{f_b(\alpha) : \alpha \in \mathbb{R}^n_{+,1}, \hat{\alpha} = \hat{\alpha}^{(b)}\} = \{f_b(\alpha^{(b)})\} = \{(f_b(\alpha^{(b)})^T, 1/(1-b), ..., 1/(1-b))^T\}$, since $\sum_{k=1}^{m} \alpha_k^{(b)} = 1$ and therefore $\alpha_k = 0, k = m+1, ..., n$. Finally, let $b \geq 1$. Due to $f_b(\hat{\alpha}^{(b)}) = (f_b(\hat{\alpha}^{(b)})^T, +\infty, ..., +\infty)^T$ and $C_{(b)} \subset \mathbb{R}^n$, we find $f_b(\hat{\alpha}^{(b)}) \notin C_{(b)}$. Suppose that z lies in $V_{(b)o}(\lambda)$. According to Lemma 13.4 ($f(0) = +\infty$) there is then an $\alpha \in ri\,\mathbb{R}^n_{+,1}$ with $z \geq f(\alpha)$ and therefore $v^{(n)*}_{(b)}(\lambda) = \lambda^T z = \hat{\lambda}^T \hat{z} = \lambda^T f_b(\alpha) = \hat{\lambda}^T f_b(\hat{\alpha})$. Because of $\alpha \in ri\,\mathbb{R}^n_{+,1}$, it is $\sum_{k=1}^{m} \alpha_k < 1$ and therefore $\hat{\alpha} \neq \hat{\alpha}^{(b)}$. On the other hand we have $\gamma = \frac{1}{2}\alpha + \frac{1}{2}\tilde{\alpha}^{(b)} \in ri\,\mathbb{R}^n_{+,1}$. This yields $v^{(n)*}_{(b)}(\lambda) \leq \lambda^T f_b(\gamma) = \hat{\lambda}^T f_b(\hat{\gamma}) < \frac{1}{2}\hat{\lambda}' f_b(\hat{\alpha}) + \frac{1}{2}\hat{\lambda}^T f_b(\hat{\alpha})^{(b)}) = v^{(n)*}_{(b)}(\lambda)$, which is again a contradiction. Consequently, $V_{(b)o}(\lambda) = \emptyset$ for $b \geq 1$ and $\lambda \notin ri\,\mathbb{R}^n_{+,1}$. □

Remark 13.8

(I) Obviously, $v^*_{(1)}(\lambda) = \sum_{k=1}^{n} \lambda_k \log \frac{1}{\lambda_k}$ is the *(Shannon-) entropy* of the discrete distribution λ.

(II) Assume that $\lambda_1 > 0, ..., \lambda_m > 0, \lambda_{m+1} = ... = \lambda_n = 0$. For $b > 0, b \neq 1$, from (13.27b) we get

$$\begin{aligned} v^*_{(b)}(\lambda) &= \frac{1}{1-b} \sum_{k=1}^{m} \lambda_k (1 - ({\lambda_k}^{1/b} / \sum_{k=1}^{m} {\lambda_k}^{1/b})^{1-b}) \\ &= \frac{1}{1-b} (1 - \sum_{k=1}^{m} {\lambda_k}^{1/b} / (\sum_{k=1}^{m} {\lambda_k}^{1/b})^{1-b}) \\ &= \frac{1}{1-b} (1 - \sum_{k=1}^{m} {\lambda_k}^{1/b})^b) = \frac{1}{1-b} (1 - (\sum_{k=1}^{n} {\lambda_k}^{1/b})^b). \end{aligned} \tag{13.30}$$

Hence,

$$v^*_{(b)}(\lambda) = \frac{1}{1-b}(1 - M_{(1/b)}(\lambda)),$$

provided that—see [6]—the *mean* $M_r(z)$, $z \in \mathbb{R}^n_+$ is defined by $M_r(z) = (\sum_{k=1}^{n} {z_k}^r)^{1/r}$. According to [6], where one finds also other properties of M_r, M_r is convex for $r > 1$ and concave for $r < 1$, which follows from the concavity of $v^*(\cdot)$.

Having $V_{(b)o}(\lambda)$, $b \geq 0$, also the functions $H^{(b)} = H^{f_b}$ and $h^{(b)} = h^{(f_b)}$ can be determined.

Corollary 13.9

(I) For $b = 0$ we have

$$H^{(o)}(\lambda, \beta) = \sup\{1 - \lambda^T \alpha : \alpha \in R^n_{+,1}, \beta^T \alpha = \max_{1 \leq k \leq n} \beta_k\} \text{ for all } \lambda, \beta \in \mathbb{R}^n_{+,1},$$

$$h^{(o)}(\lambda, \beta) = \inf\{1 - \lambda^T \alpha : \alpha \in R^n_{+,1}, \beta^T \alpha = \max_{1 \leq k \leq n} \beta_k\} \text{ for all } \lambda, \beta \in \mathbb{R}^n_{+,1},$$

(II) If $0 < b < 1$, then with $\alpha^{(b)} = ({\alpha_k}^{1/b} / \sum_{k=1}^{n} {\alpha_k}^{1/b})_{k=1,\ldots,n}$, we get

$$H^{(b)}(\lambda, \beta) = h^{(b)}(\lambda, \beta) = \sum_{k=1}^{n} \lambda_k f_b({\alpha_k}^{(b)})$$
$$= \tfrac{1}{1-b}(1 - \sum_{k=1}^{n} \lambda_k ({\beta_k}^{1/b} / \sum_{k=1}^{n} {\beta_k}^{1/b})^{1-b}) \text{ for all } \lambda, \beta \in \mathbb{R}^n_{+,1}.$$

(III) If $b = 1$, then

$$H^{(1)}(\lambda, \beta) = h^{(1)}(\lambda, \beta) = \sum_{k=1}^{n} \lambda_k \log(1/\beta_k), \lambda, \beta \in \mathbb{R}^n_{+,1} (\text{with} \log \frac{1}{0} = +\infty).$$

(IV) If $b > 1$, then

$$H^{(b)}(\lambda, \beta) = h^{(b)}(\lambda, \beta) = \lambda^T f_b(\alpha^{(b)})$$
$$= \sum_{k=1}^{n} \lambda_k \tfrac{1}{1-b}(1 - ({\beta_k}^{1/b} / \sum_{k=1}^{n} {\beta_k}^{1/b})^{1-b}), \text{ for all } \lambda, \beta \in \mathbb{R}^n_{+,1}.$$

Remark 13.9 Corresponding to Theorem 13.3, we observe that

$$H^{(1)}(\lambda, \beta) = h^{(1)}(\lambda, \beta) = \sum_{k=1}^{n} \lambda_k \log(1/\beta_k), \lambda, \beta \in \mathbb{R}^n_{+,1}$$

is the Kerridge-Inaccuracy for the hypothesis "$P_\omega = \beta$", while $P_\omega = \lambda$ is the true distribution. However, this justifies the notation *generalized inaccuracy function* for $H(\lambda, \beta)$ and $h(\lambda, \beta)$.

13.2.2 Representation of $H_\varepsilon(\lambda, \beta)$ and $H(\lambda, \beta)$ by Means of Lagrange Duality

In the following we derive a representation of $H_\varepsilon(\lambda, \beta)$ and $H(\lambda, \beta)$ which can be used also to find sufficient conditions for $H(\lambda, \beta) = H_0(\lambda, \beta)$. For this we make the following assumptions on the loss set V, cf. (13.13a), (13.13b), of the decision problem (Ω, D, v) and the probability measure λ, β on $\mathfrak{A}$: V is a convex subset of $L_1(\Omega, \mathfrak{A}, \lambda) \cap L_1(\Omega, \mathfrak{A}, \beta)$, where $-\infty < v^*(\beta) < +\infty$. Defining the mappings $F: V \to \mathbb{R}$ and $g_\varepsilon : L_1(\Omega, \mathfrak{A}, \beta) \to \mathbb{R}$, $\varepsilon \geq 0$ by

$$F(f) = \int f(\omega)\lambda(d\omega),\ f \in V$$

and

$$g_\varepsilon(f) = \int f(\omega)\beta(d\omega) - (v^*(\beta) + \varepsilon), \quad f \in L_1(\Omega, \mathfrak{A}, \beta), \varepsilon \geq 0,$$

F is an affine, real-valued functional on V and g_ε an affine, real valued functional on the linear space $X = L_1(\Omega, \mathfrak{A}, \beta)$. Moreover, it holds

$$H_\varepsilon(\lambda, \beta) = sup\{F(f) : g_\varepsilon(f) \leq 0, f \in V\},\ \varepsilon \geq 0.$$

Thus, we have to consider the following convex program in space X:

$$\min\ -F(f) \text{ s.t } g_\varepsilon(f) \leq 0, f \in V. \tag{13.31}$$

According to Luenberger [10], Sect. 8.6, Theorem 1, concerning programs of the type (13.31), we get immediately this result:

Theorem 13.4 *If $H_\varepsilon(\lambda, \beta) \in \mathbb{R}$ for $\varepsilon > 0$, then*

$$\begin{aligned} H_\varepsilon(\lambda, \beta) &= \min_{a \geq 0}(\sup_{x \in D}(v(\lambda, x) - av(\beta, x)) + a(v^*(\beta) + \varepsilon) \\ &= \min_{a \geq 0}(\sup_{x \in D}(v(\lambda, x) - a(v(\beta, x) - v^*(\beta))) + a\varepsilon), \end{aligned} \tag{13.32}$$

where the minimum in (13.32) is taken in a point $a_\varepsilon \geq 0$.

Proof The dual functional related to (13.31) is defined here by

$$\begin{aligned} \phi_\varepsilon(a) &= \inf_{f \in V}(-F(f) + ag_\varepsilon(f)) = \inf f \in V(-\int f(\omega)\lambda(d\omega) \\ &\quad + a(\int f(\omega)\beta(d\omega) - v^*(\beta) - \varepsilon)) \\ &= -(\sup_{x \in D}(v(\lambda, x) - av(\beta, x) + a(v^*(\beta) + \varepsilon)). \end{aligned}$$

Due to $v^*(\beta) > -\infty$, we also have $D_{\varepsilon/2}(\beta) \neq \emptyset$. Hence, there is $f_1 \in V$ such that $\beta f_1 \leq v^*(\beta) + \frac{\varepsilon}{2} < v^*(\beta) + \varepsilon$ and therefore $g_\varepsilon(f_1) < 0$ for all $\varepsilon > 0$. According to our assumptions, (13.31) has a finite infimum, hence, all assumptions in the above-mentioned theorem of Luenberger are fulfilled. Consequently, inf (13.31) $= \max_{a>0} \phi_\varepsilon(a),$, where the maximum is taken in a point $a_\varepsilon \geq 0$. Thus,

$$\begin{aligned} H_\varepsilon(\lambda, \beta) &= -\inf\ (13.32) = -\max_{a\geq 0} \phi_\varepsilon(a) = \min_{a\geq 0}(-\phi_\varepsilon(a)) \\ &= \min_{a\geq 0}(\sup_{x\in D}(v(\lambda, x) - av(\beta, x) + a(v^*(\beta) + \varepsilon)) \end{aligned}$$

and the maximum is taken in point $a_\varepsilon \geq 0$. □

Remark 13.10 Note that for the derivation of (13.32) only the convexity of $V = \{v(\cdot, x) : x \in D\} \subset L_1(\Omega, \mathfrak{A}, \lambda) \cap L_1(\Omega, \mathfrak{A}, \beta)$, $D_\varepsilon(\beta) \neq \emptyset$ and the condition $H_\varepsilon(\lambda, \beta) \in \mathbb{R}$ was needed.

For a comparison between $H(\lambda, \beta)$ and $H_0(\lambda, \beta)$ we show the following result:

Theorem 13.5 *Suppose that $H_{\bar\varepsilon}(\lambda, \beta) < +\infty$ for an $\bar\varepsilon > 0$. Then $H(\lambda, \beta)$ has the representation*

$$H(\lambda, \beta) = \inf_{a\in R}(\sup_{x\in D}(v(\lambda, x) - av(\beta, x)) + av^*(\beta)). \tag{13.33}$$

Proof Let h denote the right-hand side of (13.33). Then, $h \leq H(\lambda, \beta)$. Put now

$$\delta(a) = \sup_{x\in D}(v(\lambda, x) - a(v(\beta, x) - v^*(\beta))), \quad a \in \mathbb{R}.$$

If $a_1 < a_2$, then $a_1(v(\beta, x) - v^*(\beta) \leq a_2(v(\beta, x) - v^*(\beta))$, $x \in D$, since $v(\beta, x) \geq v^*(\beta)$, $x \in D$. Hence, $-a_1(v(\beta, x) - v^*(\beta)) \geq -a_2(v(\beta, x) - v^*(\beta))$, $v(\lambda, x) - a_1(v(\beta, x) - v^*(\beta)) \geq v(\lambda, x) - a_2(v(\beta, x) - v^*(\beta))$ and therefore $\delta(a_1) \geq \delta(a_2)$. Thus, δ is monotonous decreasing. However, this yields $h = \inf_{a\in\mathbb{R}} \delta(a) = \inf_{a\geq 0} \delta(a) = H(\lambda, \beta)$. □

Suppose now that $D_0(\beta) \neq \emptyset$ and $H_{\bar\varepsilon}(\lambda, \beta) < +\infty$ for an $\bar\varepsilon > 0$, hence, $H_\varepsilon(\lambda, \beta) \in \mathbb{R}$ for $0 < \varepsilon < \bar\varepsilon$. Because of $H_0(\lambda, \beta) = \sup\{F(f) : g_0(f) = 0,\ f \in V\}$, for the consideration of $H_0(\lambda, \beta)$, in stead of (13.31), we have to consider the optimization problem

$$\min \quad -F(f) \tag{13.34a}$$

$$\text{s.t.} \quad g_0(f) = 0,\ f \in V. \tag{13.34b}$$

The dual functional related to this program reads

$$\phi_0(a) = \inf_{f\in V}(-F(f) + ag_0(f)) = -\sup_{\in D}((v(\lambda, x) - av(\beta, x)) + av^*(\beta)) = -\delta(a). \tag{13.35}$$

A *Kuhn-Tucker-coefficient* related to (13.34a) is, cf. e.g., [14], Sect. 28, a value a_0, such that

$$\inf(13.34a) = \phi_0(a_0) \quad \text{or} \quad H_0(\lambda, \beta) = -\phi_0(a_0). \tag{13.36}$$

In case that such an a_0 exists, then

$$H_0(\lambda, \beta) = \inf(13.34a) = -\phi_0(a_0) = \sup_{x \in D}(v(\lambda, x) - a_0(v(\beta, x) - v^*(\beta))),$$

due to $H_0(\lambda, \beta) \leq H(\lambda, \beta)$ and (13.33), we have

$$H_0(\lambda, \beta) \leq H(\lambda, \beta) = \inf_{a \in \mathbb{R}}(\sup_{x \in D}(v(\lambda, x) - a(v(\beta, x) - v^*(\beta))))$$

$$\leq \sup_{x \in D}(v(\lambda, x) - a_0(v(\beta, x) - v^*(\beta))) = H_0(\lambda, \beta),$$

hence,

$H(\lambda, \beta) = H_0(\lambda, \beta) = \min_{a \in \mathbb{R}}(\sup_{x \in D}(v(\lambda, x) - a(v(\beta, x) - v^*(\beta)))).$

Thus, we have this result.

Theorem 13.6 *Let $D_0(\beta) \neq \emptyset$, and assume $H_{\bar{\varepsilon}}(\lambda, \beta) < +\infty$ for a certain $\bar{\varepsilon} > 0$. If the program*

$$\begin{aligned} &\max \quad \lambda f \\ &s.t. \quad \beta f = v^*(\beta), \ f \in V \end{aligned}$$

admits a Kuhn-Tucker coefficient, i.e., an $a_0 \in \mathbb{R}$, such that

$$\sup\{\lambda f : \beta f = v^*(\beta), f \in V\} = \sup\{f \in V\}(\lambda f - a_0(\beta f - v^*(\beta))),$$

then

$$H(\lambda; \beta) = H_0(\lambda, \beta) = \min_{a \in \mathbb{R}}(\sup_{x \in D}(v(\lambda, x) - a(v(\beta, x) - v^*(\beta)))), \tag{13.37}$$

and the minimum in (13.37) is taken at a point $a_0 \in \mathbb{R}$.

13.3 Generalized Divergence and Generalized Minimum Discrimination Information

13.3.1 Generalized Divergence

As in the preceding section, assume that $P_{\tilde{\omega}} = \lambda$ is the true probability distribution of $\tilde{\omega}$ and let denote $P_{\tilde{\omega}} = \beta$ a certain hypothesis on the true distribution λ; For all $\varepsilon > 0$ suppose that $D_\varepsilon \neq \emptyset$. Selecting a decision $x \in D_\varepsilon$, then with respect to the true distribution λ, with respect to the hypothesis β, resp., we have the error

$$e_1 = e_1(\lambda, x) = v(\lambda, x) - v^*(\lambda),$$

$$e_2 = e_2(\lambda, x) = v(\lambda, x) - v^*(\beta),$$

resp., where $e_1(\lambda, x)$, $e_2(\lambda, x)$, are defined only if $v^*(\lambda) \in \mathbb{R}$, $v^*(\beta) \in \mathbb{R}$, respectively.

Evaluating this error still by means of a function γ, then

$$I^e_{\gamma,\varepsilon}(\lambda, \beta) = \sup\{\gamma(e(\lambda, x)) : x \in D_\varepsilon(\beta)\}, \quad e = e_1, e_2,$$

$$J^e_{\gamma,\varepsilon}(\lambda, \beta) = \inf\{\gamma(e(\lambda, x)) : x \in D_\varepsilon(\beta)\}, \quad e = e_1, e_2$$

denotes the maximum, minimum error relative to γ, for the computation of an ε—optimal decision based on the hypothesis β, while $P_{\tilde{\omega}} = \lambda$ is the true distribution. Hence, as far as the limits under consideration exist, we define the class of generalized divergences by

$$I^e_\gamma(\lambda, \beta) = \lim_{\varepsilon \downarrow 0} I^e_{\gamma,\varepsilon}(\lambda, \beta) = \lim_{\varepsilon \downarrow 0}(\sup_{x \in D_\varepsilon(\beta)} \gamma(e(\lambda, x))), e = e_1, e_2$$

and

$$J^e_\gamma(\lambda, \beta) = \lim_{\varepsilon \downarrow 0} J^e_{\gamma,\varepsilon}(\lambda, \beta) = \lim_{\varepsilon \downarrow 0}(\inf_{x \in D_\varepsilon(\beta)} \gamma(e(\lambda, x))), e = e_1, e_2.$$

We consider now I^e, J^e, e_2 for some special cases of the cost function γ. For this purpose we set

$$I(\lambda, \beta) = H(\lambda, \beta) - H(\lambda, \lambda) = H(\lambda, \beta) - v^*(\lambda), \tag{13.38a}$$

$$J(\lambda, \beta) = h(\lambda, \beta) - h(\lambda, \lambda) = h(\lambda, \beta) - v^*(\lambda). \tag{13.38b}$$

(I) If $\gamma(t) = t$, $t \in \mathbb{R}$, then

$$I^{e_1}_{\gamma,\varepsilon}(\lambda, \beta) = \sup_{x \in D_\varepsilon(\beta)} e_1(\lambda, x) = \sup_{x \in D_\varepsilon(\beta)} (v(\lambda, x) - v^*(\lambda)) = H_\varepsilon(\lambda, \beta) - v^*(\lambda)$$

$$I^{e_2}_{\gamma,\varepsilon}(\lambda,\beta) = \sup_{x\in D_\varepsilon(\beta)} e_2(\lambda,x) = \sup_{x\in D_\varepsilon(\beta)} (v(\lambda,x) - v^*(\beta)) = H_\varepsilon(\lambda,\beta) - v^*(\lambda)$$

$$J^{e_1}_{\gamma,\varepsilon}(\lambda,\beta) = \inf_{x\in D_\varepsilon(\beta)} e_1(\lambda,x) = \inf_{x\in D_\varepsilon(\beta)} (v(\lambda,x) - v^*(\lambda)) = h_\varepsilon(\lambda,\beta) - v^*(\lambda)$$

$$J^{e_2}_{\gamma,\varepsilon}(\lambda,\beta) = \inf_{x\in D_\varepsilon(\beta)} e_2(\lambda,x) = \inf_{x\in D_\varepsilon(\beta)} (v(\lambda,x) - v^*(\beta)) = h_\varepsilon(\lambda,\beta) - v^*(\lambda).$$

Taking the limit $\varepsilon \downarrow 0$, we obtain then

$$I^{e_1}_{\gamma}(\lambda,\beta) = H(\lambda,\beta) - v^*(\lambda) = I(\lambda,\beta),$$

$$I^{e_2}_{\gamma}(\lambda,\beta) = H(\lambda,\beta) - v^*(\beta) = I(\lambda,\beta) + (H(\lambda,\lambda) - H(\beta,\beta)),$$

$$J^{e_1}_{\gamma}(\lambda,\beta) = h(\lambda,\beta) - v^*(\lambda) = J(\lambda,\beta),$$

$$J^{e_2}_{\gamma}(\lambda,\beta) = h(\lambda,\beta) - v^*(\beta) = J(\lambda,\beta) + (h(\lambda,\lambda) - h(\beta,\beta)).$$

Remark 13.11 Estimations of the variation $H(\lambda,\lambda) - H(\beta,\beta) = h(\lambda,\lambda) - h(\beta,\beta) = v^*(\lambda) - v^*(\beta)$ of the inaccuracy function $\lambda \to v^*(\lambda)$ in the transfer from λ to β can be found, e.g., in [13].

(II) Let now $\gamma(t) = |\,t|$. Because of $v(\lambda,x) \geq v^*(\lambda)$ for all $x \in D$ we have

$$I^{e_1}_{\gamma,\varepsilon}(\lambda,\beta) = \sup_{x\in D_\varepsilon(\beta)} |\,e_1(\lambda,x)\,| = \sup_{x\in D_\varepsilon(\beta)} (v(\lambda,x) - v^*(\lambda)) = H_\varepsilon(\lambda,\beta) - v^*(\lambda),$$

$$J^{e_1}_{\gamma,\varepsilon}(\lambda,\beta) = \inf_{x\in D_\varepsilon(\beta)} |\,e_1(\lambda,x)\,| = \sup_{x\in D_\varepsilon(\beta)} (v(\lambda,x) - v^*(\lambda)) = h_\varepsilon(\lambda,\beta) - v^*(\lambda).$$

Thus, also here we get

$$I^{e_1}_{\gamma}(\lambda,\beta) = H(\lambda,\beta) - v^*(\lambda) = I(\lambda,\beta),$$

$$J^{e_1}_{\gamma}(\lambda,\beta) = h(\lambda,\beta) - v^*(\lambda) = J(\lambda,\beta).$$

Furthermore, we have

$$\begin{aligned} I^{e_2}_{\gamma,\varepsilon}(\lambda,\beta) &= \sup_{x\in D_\varepsilon(\beta)} |\,e_2(\lambda,x)\,| = \max\{|\inf_{x\in D_\varepsilon(\beta)} e_2(\lambda,x)\,|, |\sup_{x\in D_\varepsilon(\beta)} e_2(\lambda,x)\,|\} \\ &= \max\{|\inf_{x\in D_\varepsilon(\beta)} (v(\lambda,x) - v^*(\beta)\,|, |\sup_{x\in D_\varepsilon(\beta)} (v(\lambda,x) - v^*(\beta))\,|\} \\ &= \max\{|\,h_\varepsilon(\lambda,\beta) - v^*(\beta)\,|, |\,H_\varepsilon(\lambda,\beta) - v^*(\beta)\,|\}. \end{aligned}$$

Because of the continuity of $(x, y) \to max\{x, y\}$, $x, y \in R$, by means of the limit $\varepsilon \downarrow 0$ we find

$$I_\gamma^{e_2}(\lambda, \beta) = \max\{|\, h(\lambda, \beta) - v^*(\beta)\, |, |\, H(\lambda, \beta) - v^*(\beta)\, |\} = \\ = \max\{|\, h(\lambda, \beta) - h(\beta, \beta)\, |, |\, H(\lambda, \beta) - H(\beta, \beta)\, |\}.$$

In the special case $H(\lambda, \beta) = h(\lambda, \beta)$, we get

$$I_\gamma^{e_2}(\lambda, \beta) = |\, H(\lambda, \beta) - H(\beta, \beta)\, |\, .$$

Especially important are the generalized divergences defined by (13.38a) and (13.38b), hence, $I(\lambda, \beta) = H(\lambda, \beta) - H(\lambda, \lambda)$ and $J(\lambda, \beta) = h(\lambda, \beta) - h(\lambda, \lambda)$. We study now I, J, where—as above—$D_\varepsilon(\beta) \neq \emptyset$, $\varepsilon > 0$ and $H(\lambda, \lambda) = h(\lambda, \lambda) = v^*(\lambda) \in \mathbb{R}$.

Corollary 13.10 *We have* $I(\lambda, \beta) \geq 0$, $J(\lambda, \beta) \geq 0$ *and* $I(\lambda, \beta) = J(\lambda, \beta) = 0$ *for* $\beta = \lambda$. *Moreover,* $I(\lambda, \beta) \geq J(\lambda, \beta)$ *und* $I(\lambda, \beta) = J(\lambda, \beta)$ *if and only if* $H(\lambda, \beta) = h(\lambda, \beta)$.

Proof According to Theorem 13.2 we have $H(\lambda, \beta) \geq h(\lambda, \beta) \geq v^*(\lambda) = h(\lambda, \lambda) = H(\lambda, \lambda)$, which yields all assertions in the above corollary. □

In order to justify the notation *generalized divergence* for the class of functions $I_\gamma^e(\lambda, \beta)$, $J_\gamma^e(\lambda, \beta)$, $e = e_1, e_2$, we consider now the case $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ with the loss set, cf. (13.23),

$$V = C_f,\ f = f_b,\ b \geq 0$$

treated in detail in the former section.

Denoting the dependence of the divergences I, J on b by $I^{(b)}$, $J^{(b)}$, then Corollary 13.9 yields immediately this result:

Corollary 13.11 *For all* $\lambda, \beta \in \mathbb{R}_{+,1}^n$ *we have*

(I)

$$I^{(0)}(\lambda, \beta) = \max_{1 \leq k \leq n} \lambda_k - inf\{\lambda^T \alpha : \alpha \in \mathbb{R}_{+,1}^n, \beta^T \alpha = \max_{1 \leq k \leq n} \beta_k\},$$

$$J^{(0)}(\lambda, \beta) = \max_{1 \leq k \leq n} \lambda_k - sup\{\lambda^T \alpha : \alpha \in \mathbb{R}_{+,1}^n, \beta^T \alpha = \max_{1 \leq k \leq n} \beta_k\};$$

(II)

$$I^{(b)}(\lambda, \beta) = J^{(b)}(\lambda, \beta) = \frac{1}{1-b} \sum_{k=1}^n \lambda_k ((\lambda_k^{1/b})^{-b} - (\beta_k^{1/b} / \sum_{k=1}^n \beta_k)^{1-b})$$

for $b > 0, b \neq 1$;

(III)

$$I^{(1)}(\lambda, \beta) = J^{(1)}(\lambda, \beta) = \sum_{k=1}^{n} \lambda_k log(\lambda_k/\beta_k).$$

Remark 13.12 Obviously, we see now that $I^{(1)}(\lambda, \beta) = J^{(1)}(\lambda, \beta)$ is the Kullback-divergence between λ and β, which justifies now the notation *generalized divergence* for I^e_γ, J^e_γ.

According to Corollary 13.10 we have $I(\lambda, \beta) = 0 \Longrightarrow J(\lambda, \beta) = 0$, and $I(\lambda, \beta) = J(\lambda, \beta) = 0$ for $\beta = \lambda$. However, $I(\lambda, \beta) = 0$ or $J(\lambda, \beta) = 0$ does not imply $\beta = \lambda$ in general.

Example 13.3 Putting $\lambda = (1/n, ..., 1/n)^T$ in the above Corollary 13.11, then for all $\beta \in \mathbb{R}^n_{+,1}$ we get

$$I^{(0)}(\lambda, \beta) = \frac{1}{n} - \inf\{\sum_{k=1}^{n} \frac{1}{n}\alpha_k : \alpha \in \mathbb{R}^n_{+,1}, \beta^T\alpha = \max_{1\le k\le n} \beta_k\} = \frac{1}{n} - \frac{1}{n} = 0,$$

$$J^{(0)}(\lambda, \beta) = \frac{1}{n} - \sup\{\sum_{k=1}^{n} \frac{1}{n}\alpha_k : \alpha \in \mathbb{R}^n_{+,1}, \beta^T\alpha = \max_{1\le k\le n} \beta_k\} = \frac{1}{n} - \frac{1}{n} = 0.$$

In this case we have therefore $\{\beta : I^{(0)}(\lambda, \beta) = 0\} = \{\beta : J^{(0)}(\lambda, \beta) = 0\} = \mathbb{R}^n_{+,1}$.

Theorem 13.7

(I) *Suppose again* $D_0(\beta) \neq \emptyset$. *If* $I(\lambda, \beta) = 0$, *then* $D_0(\lambda) \neq \emptyset$, $D_0(\beta) \subset D_0(\lambda)$, *and* $H(\lambda, \beta) = H_0(\lambda, \beta) = h_0(\lambda, \beta)$. *If in addition* $D_0(\lambda) = \{x_\lambda\}$, *then* $D_0(\beta) = \{x_\beta\}$ *with* $x_\beta = x_\lambda$.

(II) *In the case* $D_0(\beta) = \{x_\beta\}$, $H(\lambda, \beta) = v(\lambda, x_\beta)$ *for all* $\lambda, \beta \in \Lambda$ *with a subset* Λ *of* $ca_{+,1}(\Omega, \mathfrak{A})$, *then* $I(\lambda, \beta) = \lambda(v(., x_\beta) - v(., x_\lambda))$ *and* $I(\lambda, \beta) = 0 \Longleftrightarrow x_\lambda = x_\beta$, *provided that* $\lambda, \beta \in \Lambda$.

(III) *Let* $D_0(\beta) = \{x_\beta\}$ *and* $H(\lambda, \beta) = v(\lambda, x_\beta)$ *for all* $\lambda, \beta \in \Lambda$. *If* $\hat{x} \in D$ *denotes then a least element of* D *with respect to the order "*$\preceq_\Lambda$*", then* $x_\beta = \hat{x}$, $\beta \in \Lambda$, $H(\lambda, \beta) = v(\lambda, \hat{x})$ *and* $I(\lambda, \beta) = 0$, $\lambda, \beta \in \Lambda$.

The representations of $H(\lambda, \beta)$ and $h(\lambda, \beta)$ given in the Theorems 13.4 and 13.5 yield the following representation of I, J:

Corollary 13.12

(I) *If* $H_{\bar{\varepsilon}}(\lambda, \beta) < +\infty$ *for an* $\bar{\varepsilon} > 0$ *and* $v^*(\lambda) \in \mathbb{R}$, *then*

$$I(\lambda, \beta) = \inf_{a\ge 0}(\sup_{x\in D}((v(\lambda, x) - v^*(\lambda)) - a(v(\beta, x) - v^*(\beta)))));$$

(II) If $v^*(\lambda) \in \mathbb{R}$ *and* $v^*(\beta) \in \mathbb{R}$, *then*

$$J(\lambda, \beta) = \sup_{a \geq 0} \sup_{x \in D}((v(\lambda, x) - v^*(\lambda)) - a(v(\beta, x) - v^*(\beta)))).$$

Remark 13.13 Equation $I(\lambda, \beta) = 0$, $J(\lambda, \beta) = 0$. Having $I(\lambda, \beta) = v(\lambda, x_\beta) - v(\lambda, x_\lambda)$, $J(\lambda, \beta) = v(\lambda, x_\beta) - v(\lambda, x_\lambda)$, resp., with two elements $x_\lambda, x_\beta \in D$, such that $D_0(\lambda) = \{x_\lambda\}$, $D_0(\beta) = \{x_\beta\}$, then $I(\lambda, \beta) = 0$, $J(\lambda, \beta) = 0$, resp., provided that $x_\lambda = x_\beta \equiv x^0$, hence, if the true distribution λ as well as the hypothesis β yield the same (unique) optimal decision $x^0 \in D$. See also the following interpretation of I, J.

As can be seen from Corollaries 13.10, 13.11 and Theorem 13.7 the generalized divergences $I(\lambda, \beta)$, $J(\lambda, \beta)$ can be considered as measures for the deviation between the probability measures λ and β relative to the decision problem (Ω, D, v) or to the loss set V.

Based on the meaning of I and J, we introduce the following definition:

Definition 13.4

(I) The right-I-ρ-, right-J-ρ-neighborhood of a distribution $\lambda \in ca_{+,1}(\Omega, \mathfrak{A})$ with $v^*(\lambda) \in \mathbb{R}$ is the set defined by

$$U_\rho^{I,r}(\lambda)(U_\rho^{J,r}(\lambda), resp.) = \{\beta \in ca_{+,1}(\Omega, \mathfrak{A}) : D_\varepsilon(\beta) \neq \emptyset, \varepsilon > 0, I(\lambda, \beta) < \rho \text{ resp.} J(\lambda, \beta) < \rho\}$$

(II) The left-I-ρ-, left-J-ρ-neighborhood of a distribution $\beta \in ca_{+,1}(\Omega, \mathfrak{A})$ with $D_\varepsilon(\beta) \neq \emptyset, \varepsilon > 0$ is the set defined by

$$U_\rho^{I,l}(\beta)(U_\rho^{J,l}(\beta), resp.) = \{\lambda \in ca_{+,1}(\Omega, \mathfrak{A}) : v^*(\lambda) \in \mathbb{R}, I(\lambda, \beta) < \rho \text{ resp.} J(\lambda, \beta) < \rho\}.$$

Of course, the divergences I, J, yield also the notion of a "convergence":

Definition 13.5

(I) A sequence (β^j) in $ca_{+,1}(\Omega, \mathfrak{A})$ is called right-I-, right-J-convergent, resp., toward an element $\lambda \in ca_{+,1}(\Omega, \mathfrak{A})$, $\beta^j \to^{I,r} \lambda$, $\beta^j \to^{J,r} \lambda$, $j \to \infty$, resp., provided that $D_\varepsilon(\beta^j) \neq \emptyset, \varepsilon > 0$, $j = 1, 2, ...$, $v^*(\lambda) \in \mathbb{R}$ and $I(\lambda, \beta^j) \to 0$, $J(\lambda, \beta^j) \to 0$, $j \to \infty$, respectively.

(II) A sequence (λ^k) in $ca_{+,1}(\Omega, \mathfrak{A})$ is called left-I- , left-J-convergent, resp., toward an element $\beta \in ca_{+,1}(\Omega, \mathfrak{A})$, $\lambda^k \to^{I,l} \beta$, $\lambda^k \to^{J,l} \beta$, $j \to \infty$, resp, provided that $D_\varepsilon(\beta) \neq \emptyset, \varepsilon > 0$, $v^*(\lambda^k) \in \mathbb{R}$, $k = 1, 2, ...$, and $I(\lambda^k, \beta) \to 0$, $J(\lambda^k, \beta) \to 0, k \to \infty$, respectively.

Remark 13.14 The distinction between the left- and right-convergence, see the above definitions, is necessary, since $I(\lambda, \beta) \neq I(\lambda, \beta)$, $J(\lambda, \beta) \neq J(\lambda, \beta)$ in general, see the following example.

Example 13.4 Consider $\Omega = \omega_1, ..., \omega_n)$ and $V = C_{(1/2)}$. According to Corollary 13.11b we have $I^{(1/2)}(\lambda, \beta) = J^{(1/2)}(\lambda, \beta) = 2(\| \lambda \| - \frac{\lambda^T \beta}{\|\beta\|}), \lambda, \beta \in \mathbb{R}^n_{+,1}$, where $\| \cdot \|$ denotes the Euclidean norm. Then, $I^{(1/2)}(\beta, \lambda) = 2(\| \beta \| - \frac{\beta^T \lambda}{\|\lambda\|})$, and for $\lambda = (1, 0, ..., 0)^T, \beta = (1/n, ..., 1/n)^T$ we get $I^{(1/2)}(\lambda, \beta) = 2(1 - 1/n^{1/2})$ and $I^{(1/2)}(\beta, \lambda) = 2(1/n^{1/2} - 1/n) = (1/n^{1/2}) I(\lambda, \beta)$. Hence, we have $I^{(1/2)}(\beta, \lambda) \neq I^{(1/2)}(\lambda, \beta)$ for $n > 1$. We still mention that $H^{(1/2)}(\lambda, \beta) = 2(1 - \frac{\lambda^T \beta}{\|\beta\|})$, hence, $H^{(1/2)}(\beta, \lambda) = 2(1 - \frac{\beta' \lambda}{\|\lambda\|})$. Consequently, $H^{(1/2)}(\lambda, \beta) = H^{(1/2)}(\beta, \lambda)$ holds if and only if $\| \beta \| = \| \lambda \|$. However, this holds not for all $\lambda, \beta \in \mathbb{R}^n$.

For the class of inaccuracy functions $H^{(b)}(\lambda, \beta), h^{(b)}(\lambda, \beta), b \geq 0$ with $H^{(b)}(\lambda, \beta) = h^{(b)}(\lambda, \beta)$ for $b > 0$, given in Corollary 13.9, we have this result:

Corollary 13.13 *Let* $I(\lambda, \beta) = H^{(b)}(\lambda, \beta) - H^{(b)}(\lambda, \lambda), \lambda, \beta \in \mathbb{R}^n_{+,1}, b > 0$. *Then* $U^{I,l}_\rho(\beta)$ *is convex for all* $\rho > 0, \beta \in \mathbb{R}^n_{+,1}$ *and each* $b > 0$; $U^{I,r}_\rho(\lambda)$ *is convex for all* $\rho > 0, \lambda \in \mathbb{R}^n_{+,1}$ *and each* $1/2 \leq b \leq 1$.

Remark 13.15 Generalizations of the Kullback-Divergence.

Pure mathematical generalizations of the Kullback-divergence

$$I^{(l)}(\lambda, \beta) = \int p_\lambda(\omega) \log(p_\lambda(\omega)/p_\beta(\omega)) m(d\omega)$$

as well as of the related Kerridge-inaccuracy

$$H^{(l)}(\lambda, \beta) = \int p_\lambda(\omega) \log(1/p_\beta(\omega)) m(d\omega),$$

where $p_\lambda = \frac{d\lambda}{dm}$, $p_\beta = \frac{d\beta}{dm}$ and m denotes a measure on $(\Omega, \mathfrak{A})$, are suggested by several authors, see, e.g., [2, 3, 5]. We mention here, cf. [2], the f-divergence

$$J_f(\lambda, \beta) = \int p_\beta(\omega) f(p_\lambda(\omega)/p_\beta(\omega)) m(d\omega),$$

where f: $\mathbb{R}_+ \to \mathbb{R}$ is a convex function.

In these papers information can be found about the type of *geometry* induced by an f-divergence, i.e., by the related system of J_f-neighborhoods, see Definition 13.4 on (subsets of) $ca_{+,1}$. For example, in [4] is shown that certain topological properties of $J_f(\lambda, \beta)$ with $f(t) = t \log(t)$, hence, $J_f(\lambda, \beta) = I^{(l)}(\lambda, \beta))$, are related to the squared distance $d^2(\lambda, \beta)$ of the Euclidean distance $d(\lambda, \beta) = \| \lambda - \beta \|, \lambda, \beta \in H$ in a Hilbert space.

In the following we show now that also the generalized divergences I, J defined by (13.38a), (13.38b) have similar properties as the squared Euclidean distance in a Hilbert space.

13.3.2 I-, J-Projections

As was shown in the literature, see, e.g., [5, 7, 9], in the information-theoretical foundation of statistics, the following minimization problem plays a major role:

$$\begin{aligned} &\min \quad I^{(1)}(\lambda, \beta) \\ &\text{s.t.} \quad \lambda \in C. \end{aligned} \tag{13.39}$$

Here, $I^{(1)}(\lambda, \beta)$ denotes the die Kullback-divergence between a given probability measure $\beta \ll m$(measure on $\mathfrak{A}$) and $\lambda \in C$, where C is a certain subset of $\{\lambda \in ca_{+,1}(\Omega, \mathfrak{A}) : \lambda \ll m\}$.

In order to find a further relation between the divergences I, J and the squared Euclidean distance in a Hilbert space H, we replace the divergence $I^{(l)}(\lambda, \beta)$ in (13.39) by $\| \lambda - \beta \|^2$, where λ, β, C, resp., are considered as elements, a subset of a Hilbert space H, then we obtain the optimization problem

$$\begin{aligned} &\min \quad \| \lambda - \beta \|^2 \\ &\text{s.t.} \quad \lambda \in C. \end{aligned} \tag{13.40}$$

However, this problem represents the projection of $\beta \in H$ onto the subset $C \subset H$. As is well known, in case of a convex set C, a solution β_0 of (13.40) is then characterized by the condition

$$< \beta_0 - \beta, \lambda - \beta_0 > \ \geq \ 0 \quad \text{for all } \lambda \in C, \tag{13.41}$$

where $< \lambda, \beta >$ denotes the scalar product in the Hilbert space H. Putting $d^2(\lambda, \beta) = \| \lambda - \beta \|^2$, then (13.41) is equivalent with

$$d^2(\lambda, \beta) \geq d^2(\lambda, \beta_0) + d^2(\beta_0, \beta) \quad \text{for all } \lambda \in C, \tag{13.42}$$

where in (13.41) and in (13.42) the equality sign (and therefore, of course, the theorem of Pythagoras) holds, if β_0 lies in the relative algebraic interior of C. As was shown in [4], the optimization problem (13.39) can also be interpreted as a (generalized) projection problem, sind a solution β_0 (13.39) can be characterized by the condition

$$I^{(1)}(\lambda, \beta) \geq I^{(1)}(\lambda, \beta_0) + I^{(1)}(\beta_0, \beta) \quad \text{for all } \lambda \in C \tag{13.43}$$

analogous to (13.42).

We now show that a corresponding result can also be obtained for the minimization problems

$$\begin{aligned} \min \quad & I(\lambda, \beta) \\ \text{s.t.} \quad & \lambda \in C \end{aligned} \tag{13.44}$$

and

$$\begin{aligned} \min \quad & J(\lambda, \beta) \\ \text{s.t.} \quad & \lambda \in C, \end{aligned} \tag{13.45}$$

where $I(\lambda, \beta)$, $J(\lambda, \beta)$ denote the divergences according to (13.38a), (13.38b).

Let denote Λ a convex subset of $ca_{+,1}$, such that $H(\lambda, \beta)$, $h(\lambda, \beta)$ are defined and $H(\lambda, \beta) \in \mathbb{R}$, $h(\lambda, \beta) \in \mathbb{R}$, $v^*(\lambda) \in \mathbb{R}$ for all $\lambda, \beta \in \Lambda$. Moreover, let be β a fixed element of Λ and C a subset of Λ. Now, a "projection" of β onto C is defined as follows:

Definition 13.6 A solution β_0 of (13.44), (13.45), resp., is called an I-, an J-projection, resp., of β onto C.

Some properties of I-, J-projections are given in the following:

Theorem 13.8 *Suppose that C is convex, and $\lambda \to H(\lambda, \lambda_0)$, $\lambda \to h(\lambda, \lambda_0)$, resp., is affine linear on Λ for $\lambda_0 = \beta$ and all $\lambda_0 \in C$. Moreover, assume that for all $\lambda, \lambda_0 \in C$ the continuity condition $H(\lambda, \lambda_0 + t(\lambda - \lambda_0)) \to H(\lambda, \lambda_0)$, $h(\lambda, \lambda_0 + t(\lambda - \lambda_0)) \to h(\lambda, \lambda_0)$ holds for $t \downarrow 0$.*

(I) A necessary condition for an I-, J-projection β_0, resp., of β onto C is then the condition (analogous to (13.42), (13.43))

$$I(\lambda, \beta) \geq I(\lambda, \beta_0) + I(\beta_0, \beta) \quad \textit{for all} \lambda \in C, \tag{13.46}$$

$$J(\lambda, \beta) \geq J(\lambda, \beta_0) + I(\beta_0, \beta) \quad \textit{for all } \lambda \in C, \tag{13.47}$$

resp., where the sign "=" holds, provided that β_0 lies in the relative algebraic interior of C.

(II) If $\lim_{t \downarrow 0} \frac{1}{t} I(\beta_0, \beta_0 + t(\lambda - \beta_0)) = \lim_{t \downarrow 0} \frac{1}{t}(H(\beta_0, \beta_0 + t(\lambda - \beta_0)) - H(\beta_0, \beta_0)) = 0$, $\lim_{t \downarrow 0} \frac{1}{t} J(\beta_0, \beta_0 + t(\lambda - \beta_0)) = \lim_{t \downarrow 0} \frac{1}{t}(h(\beta_0, \beta_0 + t(\lambda - \beta_0)) - h(\beta_0, \beta_0)) = 0$, resp., for all $x \in C$ and a $\beta_o \in C$, then (13.46), (13.47) is also sufficient for an I-, J-projection β_0 of β onto C.

13.3.3 Minimum Discrimination Information

An important reason for the consideration of the I-, J-projections according to Definition 13.6 is the following generalization of the *minimum discrimination information*, a concept that was introduced in [9] for the foundation of methods of statistics. We suppose here the unknown (partly known) probability distribution $P_{\tilde{\omega}} = \lambda$ of $\tilde{\omega}$ lies in a subset Λ of $ca_{+,1}(\Omega, \mathfrak{A})$ and satisfies an equation of the type

$$\int T(\omega)\lambda(d\omega) = \theta(\equiv ET(\omega)), \tag{13.48}$$

where $T : \Omega \to \theta$ is a measurable mapping from $(\Omega, \mathfrak{A})$ into a further measurable set $(\Theta, \mathcal{B})$, and θ is an element of Θ. The element θ is interpreted as a certain parameter of the distribution P_ω; moreover, it is assumed that estimates $\tilde{\theta} = \tilde{\theta}_N(\omega_1, ..., \omega_n)$ are available for θ, where ω_k is a realization of $\tilde{\omega}$.

For a known parameter θ the set

$$C = C(\theta) = \{\lambda \in ca_{+,1}(\Omega, \mathfrak{A}) : \lambda \in \Lambda, \int T(\omega)\lambda(d\omega) = \theta\}$$

describes the information available on $P_{\tilde{\omega}} = \lambda$. For a given hypothesis $P_{\tilde{\omega}} = \beta$ the I-projection of β onto $C(\theta)$ describes then the nearest element of $C = C(\theta)$ to β, and

$$\begin{aligned} I(\star, \beta) = I(\star, \beta; \theta) &= \inf\{I(\lambda, \beta) : \lambda \in C(\theta)\} \\ &= \inf\{I(\lambda, \beta) : \lambda \in \Lambda, \int T(\omega)\lambda(d\omega) = \theta\} \end{aligned} \tag{13.49}$$

denotes the distance between β and $C(\theta)$ (often identified with θ). Hence, an increasing distance $I(\star, \beta)$ between β and $C(\theta)$ means a decreasing quality of the hypothesis $P_{\tilde{\omega}} = \beta$.

Corresponding to [9] we introduce therefore the following notion.

Definition 13.7 The value $I(\star, \beta) = I(\star, \beta; \theta)$ is called the minimum useful discrimination information—relative to the decision problem (Ω, D, v)—against the (zero-) hypothesis $P_{\tilde{\omega}} = \beta$.

Remark 13.16 *Useful* Discrimination Information.

The notion useful discrimination information emphasizes the fact that the generalized divergence $I(\lambda, \beta)$ measures the difference between the distributions λ and β relative to a (subsequent) decision problem (Ω, D, v); see also the definition of *economic information measures* used in [5, 11].

We show now some properties of the function $\theta \to I(\star, \beta; \theta)$.

Lemma 13.5 *Let Λ be convex, Θ a linear parameter space and $\lambda \to I(\lambda, \beta)$ convex on Λ. Furthermore, let $\Theta_0 = \{\theta \in \Theta : \text{there is } \lambda \in \Lambda, \text{ such that } \int T(\omega)\lambda(d\omega) = \theta\}$. Then Θ_0 is convex, and $\theta \to I(\star, \beta; \theta)$ is convex on Θ_0.*

Proof Let $\theta_1, \theta_2 \in \Theta_0$ and $0 < \alpha < 1$. Then there are elements $\lambda_1, \lambda_2 \in \Lambda$, such that $\theta_i = \int T(\omega)\lambda_i(d\omega), i = 1, 2.$ This yields $\int T(\omega)(\alpha\lambda_1 + (1-\alpha)\lambda_2)(d\omega) = \alpha \int T(\omega)\lambda_1(d\omega) + (1-\alpha) \int T(\omega)\lambda_2(d\omega)$ and $\alpha\lambda_1 + (1-\alpha)\lambda_2 \in \Lambda$, hence, $\alpha\theta_1 + (1-\alpha)\theta_2 \in \Theta_0$. Furthermore,

$$I(\star, \beta; \alpha\theta_1 + (1-\alpha)\theta_2 \leq I(\alpha\lambda_1 + (1-\alpha)\lambda_2, \beta) \leq \alpha I(\lambda_1, \beta) + (1-\alpha)I(\lambda_2, \beta),$$

which yields the rest of the assertion, since, up to the above conditions, λ_1, λ_2 were arbitrary. □

Remark 13.17 Convexity of $I(., \beta)$. Because of the concavity of $\lambda \to v^*(\lambda) = H(\lambda, \lambda)$, the function $\lambda \to I(\lambda, \beta)$ is convex, provided that $H(\lambda, \beta) = H_0(\lambda, \beta) = sup\{\lambda f : f \in V_0(\beta)\}$.

If Θ is a finite-dimensional space, then the convexity of $\theta \to I(\star, \beta; \theta)$ on Θ_0 yields the continuity of this function—at least—on the relative interior $ri\,\Theta_0$ of Θ_0. If the function $\lambda \to I(\star, \beta; \lambda)$ is continuous on a sufficiently large range of definition, then

$$I(\star, \beta; \hat{\theta}_N) \to I(\star, \beta; \theta) = I(\star, \beta) \quad \text{a.s.},$$

provided that $\hat{\theta}_N \to \theta$ a.s., where $\hat{\theta}_1, \hat{\theta}_2, ...$ is a sequence of estimation functions for θ. In this case we interpret then

$$\hat{I}(\star, \beta) = I(\star, \beta; \hat{\theta}_N(\omega_1, ..., \omega_N)) \quad (N = 1, 2, ...)$$

as an estimate of $I(\star, \beta)$. For the testing of hypotheses we have then, cf. [9] the following procedure:

Definition 13.8 Reject the (null-) hypothesis $P_{\tilde{\omega}} = \beta$, if $\hat{I}(\star, \beta)$ is significantly large.

For illustration of this test procedure we give still the following example:

Example 13.5 Let $\Omega = \Theta = D = \mathbb{R}$ and $v(\omega, x) = (a(\omega)x - b(\omega))^2$, where $x \in \mathbb{R}$ and $a(\cdot), b(\cdot)$ are square integrable random variables. Then,

$$I(\lambda, \beta) = \overline{a^2}^{\lambda} \left(\overline{ab}^{\lambda} / \overline{a^2}^{\lambda} - \overline{ab}^{\beta} / \overline{a^2}^{\beta} \right)^2,$$

where $\overline{a^2}^{\lambda} = \int a(\omega)^2 \lambda(d\omega)$, etc. If now $T(\omega) = a(\omega)b(\omega)$, then $I(\star, \beta) = inf \left\{ \overline{a^2}^{\lambda} \left(\overline{ab}^{\lambda} / \overline{a^2}^{\lambda} - \overline{ab}^{\beta} / \overline{a^2}^{\beta} \right)^2 : \int a(\omega)b(\omega)\lambda(d\omega) = \theta \right\} = \inf_{u>0} u(\theta/u - \overline{ab}^{\beta} / \overline{a^2}^{\beta})^2$, hence, $I(\star, \beta) = 0$, provided that $sign\theta = sign(\overline{ab}^{\beta})$ and $I(\star, \beta) = 4 \mid \theta\overline{ab}^{\beta} \mid$, if $sign\theta = -sign(\overline{ab}^{\beta})$.

References

1. Arimoto, S.: Information-theoretical considerations on estimation problems. Inf. Control **19**(3), 181–194 (1971). https://doi.org/10.1016/S0019-9958(71)90065-9
2. Csiszar, I.: Information-typ measures of difference of probability distributions and indirect observations. Studia Scientiarum Hungarica **2**, 299–318 (1967)
3. Csiszar, I.: On topological properties of f-divergences. Studia Scientiarum Hungarica **2**, 329–339 (1967)
4. Csiszar, I.: I-divergence geometry of probability distributions and minimization problems. Ann. Probab. **3**(1), 146–158 (1975). https://doi.org/10.1214/aop/1176996454
5. Good, I.: What is the use of a distribution? In: P. Krishnaiah (ed.) Multivariate Analysis, Proceedings of the 2nd International Symposium on Multivariate Analysis, pp. 183–203. Academic Press, New York (1969)
6. Hardy, G., Littlewood, J., Pólya, G.: Inequalities. Cambridge University Press, London (1973)
7. Jaynes, E.: Prior probabilities. IEEE Trans. Syst. Sci. Cybern. **4**(3), 227–241 (1968). https://doi.org/10.1109/TSSC.1968.300117
8. Kerridge, D.: Inaccuracy and inference. J. R. Stat. Soc. Ser. B (Methodol.) **23**(1), 184–194 (1961). http://www.jstor.org/stable/2983856
9. Kullback, S.: Information Theory and Statistics. Wiley, New York (1959)
10. Luenberger, D.: Optimization by Vector Space Methods. Wiley, New York (1969)
11. Marschak, J.: Economics of information systems. J. Am. Stat. Assoc. **66**(333), 192–219 (1971). http://www.jstor.org/stable/2284873
12. Martino, J.: Technological Forecasting for Decision Making, 3rd edn. McGraw-Hill, New York (1993)
13. Perez, A.: Information-theoretic risk estimates in statistical decision. Kybernetika **3**, 1–21 (1967)
14. Rockafellar, R.: Convex Analysis. University Press, Princeton (1970)
15. Theil, H.: Optimal Decision Rules for Government and Industry. North-Holland, Amsterdam (1968)

Index

K. Marti, *Stochastic Optimization Methods*,
https://doi.org/10.1007/978-3-031-40059-9