Stochastic Programming Models for Optical Fiber Manufacturing

Csaba István Fábián

A dissertation submitted for the degree of
Doctor of Philosophy, Operations Research

Written under the direction of Prof. András Prékopa

Doctoral School in
Operations Research, Applied Mathematics, and Statistics
Loránd Eötvös University

Budapest, 1999

Acknowledgements

I wish to express my thanks to Prof. András Prékopa for presenting me the problem and for his constant help throughout the project: in the formulation of the model, in finding appropriate solution methods, and in the presentation of the results. I must also thank Dr. Tamás Szántai for suggesting improvements in a constructive way, and for his help in the presentation of the results. I am grateful to the Department of Operations Research, Loránd Eötvös University for enabling me to work on this dissertation.

The work was supported by the Bolyai János Foundation of the Hungarian Academy of Sciences; and by OTKA, Hungarian National Fund for Scientific Research, project no. T014102.

Contents

Chapter 1

Introduction

Optical fibers are made of solid rods of glass called preforms. The ends of the preforms are heated and fibers are drawn from them. The fibers may randomly break during the process. The random lengths of fibers produced will be referred to as semi-finished products throughout this paper. We assume that all preforms are of the same size. Preforms are processed continuously at a constant rate.

The products are standard-length fibers, manufactured by cutting semi-finished products. We make the following assumptions:

– Semi-finished products are cut into standard lengths immediately after their production.

– The number of different products (i.e., the different standard lengths) is small as compared to the numbers of semi-finished products produced in any length.

Time is divided into periods. Demand for each product is assumed to be known for each period, and deliveries are made at the ends of periods. We assume further:

– A fixed large number of preforms are to be processed during each production period.

– The way of cutting semi-finished products is determined by a cutting rule that has to be specified beforehand, and left unchanged during the production period.

Our aim is to find such a cutting rule that minimizes the expected loss due to unsatisfied demand.

The above problem is closely related to the one presented by Murr and Prékopa (1996). They distinguish between fibers on the basis of length, as well as of additional physical characteristics. In their approach the intensity of production is also a decision variable, and the aim is to minimize production costs such that all demands are satisfied with a high probability. They propose a rolling horizon model system, where each model encompasses two production periods. Each model contains a probabilistic constraint in a joint constraint form and penalizes the violations of the stochastic constraints.

If continuous probability distributions are considered, then the resulting problems can be solved by the method developed by Szántai (1988). For the case when the variables are

discrete, Prékopa (1990) developed a dual type method. Improvements and technical details of the dual type method, and implementation issues can be found in Fábián, Prékopa, and Ruf-Fiedler (1999).

The present models consider only the lengths of the fibers, and disregard other physical characteristics. However, they allow for more flexibility in cutting in the following respects:

– They do not require the selection of all suitable cutting patterns before optimization. (A cutting pattern means a way a certain length of fiber can be cut into standard lengths.)

– They use a new cutting concept, the cutting rule. A cutting rule determines the relative frequencies with which the different cutting patterns are to be used. The decision variables in the present models represent a cutting rule.

To find optimal cutting rules, stochastic programming problems are formulated and solved. Comprehensive treatment of stochastic programming problems can be found in Prékopa (1995). The solution methods presented here exploit the special structures of the cutting rules. (Cutting rules are formally described by matrices.) We show that they form a convex polyhedron. This polyhedron is not known explicitly, but can be explored step by step by the column generation method of Gilmore and Gomory (1961).

In Chapter 2 we describe the nature of randomness in fiber production, and define the concept of cutting rules.

In Chapter 3 a static model is described that takes into consideration only one production period. In Chapter 4 we present reliability considerations for the static model. The resulting problems have convex objective functions. The feasible domains are not known explicitly, but can be explored step by step. Thus the problems can be solved by extended variants of feasible direction methods. (We propose variants of the procedure P2 of Zoutendijk (1960).) A feasible direction method may take a large number of steps to approach optimal solution with prescribed precision. In the one-period case, however, a single step requires little computational effort. Details of the static model are described in Fábián (1998).

In Chapter 5 we formulate a dynamic model that takes two production periods into consideration. Under reasonable assumptions, the resulting problem has a convex objective function. The model has a special feature: first-period events affect second-period events only through the inventory of the products in stock at the beginning of the second period. The inventory can be described by a relatively small amount of data. This feature enables us to evaluate the objective function and its subgradient by simulation. In Chapters 6 and 7 we define stochastic substitutes for supporting hyperplanes, and present an evaluation method.

Finding a good estimate of the objective value and subgradient requires a relatively big effort. Hence these data are worth retaining and reusing. A feasible direction method discards these data once it moved to a new point. The classical cutting-plane method does accumulate all the information obtained, but it is unstable. Bundle methods are a refinement of the cutting-plane method. They proved stable and effective, and for this reason are obtaining increasing attention in convex optimization.

Recently, Lemaréchal, Nemirovskii, and Nesterov (1995) proposed new variants of the bundle methods. In Chapter 8 we develop a stochastic version of their Level Method. At

any stage, good upper and lower bounds for the optimal objective value are available. We exploit this feature. At the beginning, only rough estimates of the objective value and gradient are needed. As the optimum is gradually approached, more and more accurate data are computed. Such heuristics are already employed in the code of Szántai (1988). The present procedure has the advantage that the required accuracy is known at each step, hence better coordination of efforts is possible. We can keep balance between the amount of work invested into estimating the objective value and gradient on the one hand, and into the optimization method on the other hand. Moreover, the present procedure inherits stability from bundle-type methods.

It is our conviction that similar methods can be developed for general two-stage stochastic programming problems, and for problems with probabilistic constraints. More research must be done to verify it.

Chapter 2

Background

2.1 The Random Yield

The collection of semi-finished products produced from a given number of preforms will be called *yield*.

Let N designate the number of preforms that can be processed during the current production period. The material of the preforms may contain a finite number of defects (microscopic grains) that cause breakage when the fiber is drawn. Assume that the material is homogeneous in the following sense:

– equal-length intervals of fiber have the same probability distribution regarding the numbers of defects contained in them;

– the probability that a positive-length fiber contains at least one defect is positive;

– the numbers of defects in disjoint intervals are independent.

The above assumptions imply that the locations of the defects in the fiber form a homogeneous Poisson process with parameter $\lambda > 0$.

We characterize the yield of a single preform by the numbers of semi-finished products of different lengths. Truncate the length of each semi-finished product to an integer multiple of a unit length. Let H be the total length of fiber that can be drawn from a single preform, and h an integer satisfying $1 \leq h \leq \lfloor H \rfloor$. Let φ_h designate the number of those semi-finished products the lengths of which fall into the interval $[\, h,\, h+1\,)$. Such fibers will be referred to as of length h. The random yield is characterized by the vector

$$\boldsymbol{\varphi} \;=\; \Big(\, \varphi_1 \,, \ldots, \, \varphi_{\lfloor H \rfloor} \,\Big)^T \quad .$$

Proposition 1 *The random variables* $\varphi_h \;\; (h = 1, \ldots, \lfloor H \rfloor)$ *are linearly independent.*

Proof. Suppose that

$$z_0 + \sum_{h=1}^{\lfloor H \rfloor} z_h \, \varphi_h = 0$$

holds with probability 1 for some real numbers z_h $(h = 0, \ldots, \lfloor H \rfloor)$.

Since $\varphi_h = 0$ $(h = 1, \ldots, \lfloor H \rfloor)$ may occur with a positive probability, it follows that $z_0 = 0$.

For any $h = 1, \ldots, \lfloor H \rfloor$, the event $\varphi_h = 1$, and $\varphi_j = 0$ $(j = 1, \ldots, \lfloor H \rfloor;\ j \neq h)$ also has a positive probability, hence $z_h = 0$ $(h = 1, \ldots, \lfloor H \rfloor)$. This implies the assertion. □

From the linear independence it follows that the covariance matrix of the random vector $\boldsymbol{\varphi}$. is positive definite.

Now, consider the random yield of N preforms, i.e., the sum

$$\boldsymbol{\phi} = \boldsymbol{\varphi}^1 + \cdots + \boldsymbol{\varphi}^N \ , \tag{2.1}$$

where the random vector $\boldsymbol{\varphi}^i$ represents the yield of the ith preform. The random vectors $\boldsymbol{\varphi}^i$ $(i = 1, \ldots, n)$ are mutually independent and have the same distribution as $\boldsymbol{\varphi}$. A multidimensional central limit theorem states that the asymptotic distribution of the random vector (2.1) is the $\lfloor H \rfloor$-variate normal distribution. (Proof of that theorem can be found in Section 4.2.3 of Anderson (1958) or in Section 9.2 of Wilks (1962).) We use this normal approximation in Section 3.3.

2.1.1 Expectation and Variance of the Random Yield

Let $0 < q < 1$ designate the probability that a certain unit-length interval of the fiber does not contain defects. Let us consider the yield of a single preform. For a real number $0 < h \leq H$, let ξ_h designate the number of semi-finished products the lengths of which are at least h.

Moments of this random variable are computed in Fábián (1999). We have

$$\mathrm{E}\,(\xi_h) = q^h - q^h\,[H - h]\,\ln q \ ,$$

$$\begin{aligned} \mathrm{E}\,(\,\xi_h\,\xi_g\,) &= -\ 2\,q^{h+g}\,[H - h - g]_+\,\ln q + q^{h+g}\,[H - h - g]_+^2\,\ln^2 q \\ &\quad +\ \mathrm{E}\left(\xi_{\max(h,g)}\right) \ , \end{aligned}$$

where $[\,.\,]_+$ means the positive part of a real number.

The expectation vector and covariance matrix of the $\lfloor H \rfloor$-dimensional random vector $\boldsymbol{\varphi}$ can be computed from the above moments, because we have

$$\begin{aligned} \varphi_h &= \xi_h - \xi_{h+1} \qquad (h = 1, \ldots, \lfloor H \rfloor - 1) \ , \\ \varphi_{\lfloor H \rfloor} &= \xi_{\lfloor H \rfloor} \ . \end{aligned}$$

2.2 Cutting Rules

Cutting rules determine the way semi-finished product fibers are to be cut into products.

Let $h_1 < \ldots < h_m$ designate the standard product lengths that we assume to be integers. Let h be a semi-finished product length, i.e., an integer between 1 and $\lfloor H \rfloor$. A cutting pattern for length h is an m-vector $\boldsymbol{a}_h = (a_1, \ldots, a_m)^T$ whose components are non-negative integers satisfying the inequality

$$\sum_{k=1}^{m} a_k h_k \leq h \ . \tag{2.2}$$

This cutting pattern means that we cut a_k pieces of length h_k $(k = 1, \ldots, m)$. Let

$$\boldsymbol{a}_h^1 \ , \ \ldots \ , \ \boldsymbol{a}_h^{r_h}$$

designate the cutting patterns for length h. We consider convex combinations of these cutting patterns. The motivation is the following. Suppose e.g., that semi-finished products of length h are cut according to the following instructions:

> Every fourth piece is cut according to pattern $\boldsymbol{a}_h^1$,
> and all other pieces according to pattern $\boldsymbol{a}_h^2$.

In that case, the relative frequencies of the two patterns will be approximately $\frac{1}{4}$ and $\frac{3}{4}$, respectively, provided the number of semi-finished products is large enough. *The respective numbers of different products cut from semi-finished products of length h will then be proportional to the components of the m-vector*

$$\boldsymbol{a}_h = \frac{1}{4}\boldsymbol{a}_h^1 + \frac{3}{4}\boldsymbol{a}_h^2 \ .$$

Convex combinations of cutting patterns for length h will be called generalized cutting patterns for length h. Let P_h designate the set of generalized cutting patterns for length h, i.e., the convex hull of the cutting patterns for length h.

A generalized cutting pattern $\boldsymbol{a}_h \in P_h$ represents the possibility of cutting semi-finished products of length h in such a way that *the respective numbers of different products will be approximately proportional to the components of the vector* $\boldsymbol{a}_h$. In Section 2.2.1 we show how to find such cutting instructions which realize a generalized cutting pattern.

We will optimize among the generalized cutting patterns, and find the appropriate cutting instructions afterwards.

Choosing a generalized cutting pattern for each semi-finished product length means picking an element of the direct product

$$P = P_1 \times \cdots \times P_{\lfloor H \rfloor} \ .$$

We define a *cutting rule* as an element of P. Such elements are $m \times \lfloor H \rfloor$-matrices the columns of which are generalized cutting patterns for different semi-finished product lengths.

Let $A \in P$ be a cutting pattern, and $\boldsymbol{a}^k$ the kth row of A. Suppose the yield $\boldsymbol{\phi}$ is cut according to rule A. The number of products of length h_k will then be approximately $\boldsymbol{a}^k\,\boldsymbol{\phi}$.

2.2.1 Finding Cutting Instructions that Realize a Given Rule

Let $A \in P$ be a cutting rule. Consider $\boldsymbol{a}_h \in P_h$, the hth column of the matrix A . We sketch a method for finding a realization of the generalized cutting pattern $\boldsymbol{a}_h$.

Semi-finished products of length h are cut into standard lengths immediately after their production. We will look for the cutting instructions in the following form:

Bunch the semi-finished products into T-element sets, according to the order of their production;
out of each set, t^i pieces are to be cut according to pattern $\boldsymbol{a}_h^i \quad (i = 1, \ldots, r_h)$, where $t^1, \ldots, t^{r_h}$ are non-negative integers and $T = t^1 + \ldots + t^{r_h}$. (2.3)

The number of the semi-finished products to be cut, i.e. ϕ_h , may not be a multiple of T. If, however, T is small as compared to ϕ_h , then the respective relative frequencies of the different cutting patterns will be proportional to the numbers $\frac{t^i}{T}$ $(i = 1, \ldots, r_h)$. Hence *the respective numbers of different products will be approximately proportional to the components of the vector*

$$\frac{t^1}{T}\,\boldsymbol{a}_h^1 \;+\; \cdots \;+\; \frac{t^{r_h}}{T}\,\boldsymbol{a}_h^{r_h} \quad .$$

We need such non-negative integers $t^i \quad (i = 1, \ldots, r_h)$ that the above vector is a good approximation of $\boldsymbol{a}_h$.

From Carathéodory's theorem it follows that there exists a convex combination

$$\boldsymbol{a}_h \;=\; \lambda^1\,\boldsymbol{a}_h^1 \;+\; \cdots \;+\; \lambda^{r_h}\,\boldsymbol{a}_h^{r_h}$$

in which at most $m + 1$ of the multipliers λ^i are positive, where m is the number of the components of $\boldsymbol{a}_h$, i.e., the number of different products. It is easy to find such a convex combination in a constructive manner.

In order to find the appropriate cutting instruction in the form of (2.3), we need such integers $t^i \quad (i = 1, \ldots, r_h)$, whose sum is small as compared to ϕ_h , while the fractions $\frac{t^i}{T}$ approximate the multipliers λ^i . Such integers can be found, since m is small as compared to the number of semi-finished products of any length. (A quick method is to fix T first, and then choose t^i by rounding $\lambda^i T$.)

Chapter 3

A Static Model

In this chapter we take into consideration only one production preiod. The decision/observation scheme in the present case is the following:

– Decision on the cutting rule to be used in the production period.

– Observation of the random yield, i.e., the set of semi-finished products produced during the current period. (The cutting rule and the yield together determine the numbers of different products produced during the period.)

– Decision on selling products. This final decision is simple: in each product, as many pieces of fiber are sold as the market can take. (In the present model we ignore the possibility of storing products for future sale.)

Becuase of the simplicity of the second decision, the model can be considered *static*. See Chapters 8 and 9 of Prékopa (1995).

We are going to prove that the resulting problem has a strictly convex and continuously differentiable objective function. This function has to be minimized over the convex polyhedron of the cutting rules. For that purpose we use a feasible direction method, procedure P2 of Zoutendijk (1960) which, under the present conditions, assures convergence to the optimal solution. Procedure P2 is extended by the method of Gilmore and Gomory. A feasible direction method may take a large number of steps to approach optimal solution with prescribed precision. In the present case, however, a single step requires little computational effort, because the objective function and its gradient can be computed by the integration of the univariate normal probability density function. The direction and step-length finding problems can be reduced to a series of knapsack problems.

The optimization problem is formulated in Section 3.1. In Section 3.2 the mathematical properties of the objective function are derived. In Section 3.3 we present a simple procedure to compute the value and the gradient of the objective function. In Section 3.4 the algorithmic solution of the problem is outlined.

3.1 Formulation of the Problem

Let b_k designate the demand for products of length h_k $(k = 1, \ldots, m)$. For simplicity we assume $b_k > 0$ for all k. Let us form the vector $\boldsymbol{b} = (b_1, \ldots, b_m)^T$.

For a given cutting rule $A \in P$, and yield $\boldsymbol{\phi}$, the components of the vector

$$[\, \boldsymbol{b} \, - \, A\boldsymbol{\phi} \,]_+$$

are the unsatisfied demands for the different products. Let

$$\delta\,(A,\, \boldsymbol{\phi}\,) \;\; = \;\; \boldsymbol{q}^T \; [\, \boldsymbol{b} \, - \, A\boldsymbol{\phi} \,]_+$$

be the loss (unrealized profit) due to unsatisfied demand, where $\boldsymbol{q} \geq \boldsymbol{0}$ is a given cost vector, and let

$$\Delta\,(A) \;\; = \;\; \mathrm{E}\Big(\, \delta\,(A,\, \boldsymbol{\phi}\,) \,\Big) \;\; .$$

The problem is to choose a cutting rule that minimizes the expected loss over the convex polyhedron of the cutting rules, i.e.,

$$\begin{array}{ll} \min & \Delta\,(A) \\ \text{subject to} & A \in P. \end{array} \tag{3.1}$$

3.2 Properties of the Objective Function

Let us consider the loss as the sum of the losses in different products. By introducing the notation

$$\delta_k\Big(\, \boldsymbol{a}^k,\, \boldsymbol{\phi}\,\Big) \;\; = \;\; q_k \;\Big[\, b_k \, - \, \boldsymbol{a}^k\boldsymbol{\phi}\,\Big]_+ \qquad \text{and} \qquad \Delta_k\Big(\, \boldsymbol{a}^k\,\Big) \;\; = \;\; \mathrm{E}\Big(\, \delta_k\Big(\, \boldsymbol{a}^k,\, \boldsymbol{\phi}\,\Big)\,\Big)\,,$$

we have

$$\delta\,(A,\, \boldsymbol{\phi}\,) \;\; = \;\; \sum_{k=1}^{m} \;\; \delta_k\Big(\, \boldsymbol{a}^k,\, \boldsymbol{\phi}\,\Big) \qquad \text{and} \qquad \Delta\,(A) \;\; = \;\; \sum_{k=1}^{m} \;\; \Delta_k\Big(\, \boldsymbol{a}^k\,\Big) \;\; .$$

Let us extend the definitions of δ_k and Δ_k for arbitrary m-vectors, and replace $\boldsymbol{a}_k$ with $\boldsymbol{p}$.

We investigate convexity and differentiability of Δ_k. (The most useful properties can be verified simply, without exploiting that the random vector $\boldsymbol{\phi}$ has a non-degenerate normal distribution.)

Proposition 2 *For any k, the function Δ_k is a convex function of the variable $\boldsymbol{p}$.*

Proof. The proof of convexity is simple. Let

$$\boldsymbol{p} \;\; = \;\; \rho_1\, \boldsymbol{p}_1 \;\; + \;\; \rho_2\, \boldsymbol{p}_2$$

be a convex combination of the points $\boldsymbol{p}_1, \boldsymbol{p}_2 \in \mathbb{R}^{\lfloor H \rfloor}$. The function $\delta_k(\boldsymbol{p}, \boldsymbol{\phi})$ is convex in $\boldsymbol{p}$, since it is the convex function of a linear function. Thus, for any $\boldsymbol{\phi}$, it satisfies the following inequality

$$\delta_k(\boldsymbol{p}, \boldsymbol{\phi}) \;\leq\; \rho_1\ \delta_k(\boldsymbol{p}_1, \boldsymbol{\phi}) \;+\; \rho_2\ \delta_k(\boldsymbol{p}_2, \boldsymbol{\phi}) \quad . \tag{3.2}$$

Taking expected values of both sides, we get the convexity inequality for the function Δ_k. □

Even strict convexity can be proved, because the random vector $\boldsymbol{\phi}$ has continuous probability distribution and its density function is positive in the positive orthant. Hence the expressions $b_k - \boldsymbol{p}_1{}^T\bar{\boldsymbol{\phi}}$ and $b_k - \boldsymbol{p}_2{}^T\bar{\boldsymbol{\phi}}$ have different signs with positive probability, which implies that the inequality in (3.2) is strict with positive probability.

Since the function $\delta_k(\boldsymbol{p}, \boldsymbol{\phi})$ is convex in $\boldsymbol{p}$, it is also subdifferentiable. For any $\boldsymbol{p} \in \mathbb{R}^{\lfloor H \rfloor}$ let us pick a subgradient of the above function, and denote it by $\boldsymbol{\gamma}_k(\boldsymbol{p}, \boldsymbol{\phi})$.

For a fixed $\boldsymbol{p}$, the function $\boldsymbol{\gamma}_k(\boldsymbol{p}, \boldsymbol{\phi})$ is a random vector. Denote the expectation vector by $\boldsymbol{\Gamma}_k(\boldsymbol{p})$.

Proposition 3 *Assume that for a given $\boldsymbol{p} \in \mathbb{R}^{\lfloor H \rfloor}$, the expectation vector $\boldsymbol{\Gamma}_k(\boldsymbol{p})$ exists.*

Then $\boldsymbol{\Gamma}_k(\boldsymbol{p})$ is a subgradient of the function $\Delta_k(.)$ at $\boldsymbol{p}$. Formally, the following inequality holds

$$\Delta_k(\boldsymbol{q}) \;\geq\; \Delta_k(\boldsymbol{p}) \;+\; \left[\,\boldsymbol{\Gamma}_k(\boldsymbol{p})\,\right]^T (\,\boldsymbol{q} \,-\, \boldsymbol{p}\,) \tag{3.3}$$

for any vectors $\boldsymbol{p}$ and $\boldsymbol{q}$.

Proof. The vector $\boldsymbol{\gamma}_k(\boldsymbol{p}, \boldsymbol{\phi})$ is a subgradient of the function $\delta_k(.\,, \boldsymbol{\phi})$ at $\boldsymbol{p}$. Formally, we have

$$\delta_k(\boldsymbol{q}, \boldsymbol{\phi}) \;\geq\; \delta_k(\boldsymbol{p}, \boldsymbol{\phi}) \;+\; \left[\,\boldsymbol{\gamma}_k(\boldsymbol{p}, \boldsymbol{\phi})\,\right]^T (\,\boldsymbol{q} \,-\, \boldsymbol{p}\,)$$

for any realizations of the random vector $\boldsymbol{\phi}$, and for any vectors $\boldsymbol{p}$ and $\boldsymbol{q}$. Taking expected values of both sides, we get (3.3). □

Proposition 4 *Assume that the expectation vector $\boldsymbol{\Gamma}_k(\boldsymbol{p})$ exists for every $\boldsymbol{p}$, and is a continuous function of $\boldsymbol{p}$.*

Then Δ_k is differentiable, and at $\boldsymbol{p}$, its gradient is $\boldsymbol{\Gamma}_k(\boldsymbol{p})$.

Proof. We prove that Δ_k has a unique subgradient at any point $\boldsymbol{p}$. Suppose that the vector $\boldsymbol{g}$ is a subgradient at $\boldsymbol{p}$, that is, the following inequality holds for any $\boldsymbol{q}$

$$\Delta_k(\boldsymbol{q}) \;\geq\; \Delta_k(\boldsymbol{p}) \;+\; \boldsymbol{g}^T\ (\boldsymbol{q} \,-\, \boldsymbol{p})\ .$$

$\boldsymbol{\Gamma}_k(\boldsymbol{q})$ being subgradient at $\boldsymbol{q}$, we have

$$\Delta_k(\boldsymbol{p}) \;\geq\; \Delta_k(\boldsymbol{q}) \;+\; \left[\,\boldsymbol{\Gamma}_k(\boldsymbol{q})\,\right]^T (\,\boldsymbol{p} \,-\, \boldsymbol{q}\,)\ .$$

Adding the above two inequalities, we get that

$$\left[\boldsymbol{g} - \boldsymbol{\Gamma}_k(\boldsymbol{q})\right]^T (\boldsymbol{p} - \boldsymbol{q}) \geq 0 \tag{3.4}$$

holds for any $\boldsymbol{q}$.

From continuity of the function $\boldsymbol{\Gamma}_k(\,.\,)$, it follows that $\boldsymbol{g} = \boldsymbol{\Gamma}_k(\boldsymbol{p})$. (Let $\boldsymbol{d}$ be an arbitrary direction, and let $\boldsymbol{q}_\ell = \boldsymbol{p} + \frac{1}{\ell}\boldsymbol{d}$ $(\ell = 1, 2, \ldots)$. Substitute the vectors $\boldsymbol{q}_\ell$ for $\boldsymbol{q}$ in (3.4).)

Uniqueness of the subgradient implies differentiability (see Theorem 25.1 in Rockafellar (1970)). □

Now we show that the assumption of Proposition 4 hold. For a fixed yield $\boldsymbol{\phi}$, the function $\delta_k(\boldsymbol{p}, \boldsymbol{\phi})$ is differentiable in $\boldsymbol{p}$, if $b_k \neq \boldsymbol{p}^T\boldsymbol{\phi}$ holds. The gradient can be computed as

$$\boldsymbol{\gamma}_k(\boldsymbol{p}, \boldsymbol{\phi}) = \begin{cases} \mathbf{0} \in \mathrm{I\!R}^{\lfloor H \rfloor}, & \text{if } b_k < \boldsymbol{p}^T\boldsymbol{\phi} \\ \\ -q_k\,\boldsymbol{\phi}, & \text{if } b_k > \boldsymbol{p}^T\boldsymbol{\phi}\,. \end{cases} \tag{3.5}$$

Since $\boldsymbol{\phi}$ has a non-degenerate continuous random distribution, the event $b_k = \boldsymbol{p}^T\boldsymbol{\phi}$ has 0 probability for any vector $\boldsymbol{p}$ (because we have $b_k > 0$). Thus the subgradient $\boldsymbol{\gamma}_k(\boldsymbol{p}, \boldsymbol{\phi})$ is also a gradient with probability 1, and can be computed from (3.5).

Let f designate the density function of the non-degenerate normal distribution of $\boldsymbol{\phi}$. The hth component of the expectation vector $\boldsymbol{\Gamma}_k(\boldsymbol{p})$ can be computed as

$$-q_k \int\limits_{\boldsymbol{p}^T\boldsymbol{z} \leq b_k} z_h\, f(\boldsymbol{z})\, d\boldsymbol{z}\ .$$

It is easy to verify that the above expectation exists, and is a continuous function of $\boldsymbol{p}$ (because we have $b_k > 0$).

3.3 Computation of the Value and the Gradient of the Objective Function

In this section we assume that the random vector $\boldsymbol{\phi}$ has a non-degenerate $\lfloor H \rfloor$-variate normal distribution with expectation vector $\boldsymbol{\mu}$ and covariance matrix C.

We present methods to compute the values and the gradients of the functions (of the variable $\boldsymbol{p} \in \mathrm{I\!R}^{\lfloor H \rfloor}$):

$$\Delta_k(\boldsymbol{p}) = \mathrm{E}\Big(q_k \left[b_k - \boldsymbol{p}^T\boldsymbol{\phi}\right]_+\Big) \qquad (k = 1, \ldots, m).$$

The objective function is the sum of them.

For a fixed $\boldsymbol{p} \in \mathrm{I\!R}^{\lfloor H \rfloor}$, $\boldsymbol{p}^T\boldsymbol{\phi}$ has a normal distribution with

$$\mu = \mathrm{E}(\boldsymbol{p}^T\boldsymbol{\phi}) = \boldsymbol{p}^T\boldsymbol{\mu}, \quad \sigma^2 = \mathrm{Var}(\boldsymbol{p}^T\boldsymbol{\phi}) = \boldsymbol{p}^T C\boldsymbol{p}.$$

Assuming $\boldsymbol{p} \neq \mathbf{0}$, the probability density function of $\boldsymbol{p}^T\boldsymbol{\phi}$ is

$$g(y) \;=\; \frac{1}{\sqrt{2\pi}\;\sigma}\;\; e^{-\frac{1}{2\sigma^2}\,(y-\mu)^2}, \qquad -\infty < y < \infty\;.$$

We have the equality

$$\begin{aligned}
\Delta_k\,(\boldsymbol{p}) \;&=\; q_k \int\limits_{-\infty}^{b_k} (b_k - y)\; g(y)\; dy \;=\; q_k \int\limits_{-\infty}^{b_k} (b_k - \mu)\; g(y)\; dy \;-\; q_k \int\limits_{-\infty}^{b_k} (y - \mu)\; g(y)\; dy \\
&=\; q_k\,(b_k - \mu) \int\limits_{-\infty}^{b_k} g(y)\; dy \;+\; q_k\,\frac{\sigma}{\sqrt{2\pi}} \int\limits_{-\infty}^{b_k} -\frac{1}{\sigma^2}\,(y-\mu)\; e^{-\frac{1}{2\sigma^2}\,(y-\mu)^2}\; dy \\
&=\; q_k\,(b_k - \mu)\,\Phi\left(\tfrac{b_k-\mu}{\sigma}\right) \;+\; q_k\,\tfrac{\sigma}{\sqrt{2\pi}}\; e^{-\frac{1}{2\sigma^2}\,(b_k-\mu)^2}\;\;,
\end{aligned}$$

where Φ is the distribution function of the $N(0,1)$-distribution. Most numeric program libraries provide routines to compute the values of this function. The second term can be written in the closed form Thus we have a simple method to compute the values of the objective function.

As regards the gradient of the objective function, it can be computed from the above formulae, using elementary differentiation rules. We present a different approach, though. Using Proposition 4, the gradient can be computed as sum of the following expectations:

$$\boldsymbol{\Gamma}_k\,(\,\boldsymbol{p}\,) \;=\; \mathrm{E}\Big(\,\boldsymbol{\gamma}_k\,(\,\boldsymbol{p},\,\boldsymbol{\phi}\,)\,\Big) \qquad (k = 1, \ldots, m).$$

In view of (3.5), the hth component of $\boldsymbol{\Gamma}_k\,(\,\boldsymbol{p}\,)$ is

$$-q_k\;\mathrm{E}\,\Big(\;\phi_h \;\mid\; \boldsymbol{p}^T\boldsymbol{\phi} \;<\; b_k\,\Big)\;\;\mathrm{P}\,\Big(\;\boldsymbol{p}^T\boldsymbol{\phi} \;<\; b_k\,\Big) \qquad\qquad (h = 1, \ldots, \lfloor H \rfloor)\;. \tag{3.6}$$

Let us introduce the notation

$$\boldsymbol{\psi} \;=\; \begin{pmatrix} \phi_h \\ \\ \boldsymbol{p}^T\boldsymbol{\phi} \end{pmatrix} \;=\; B\boldsymbol{\phi}, \qquad \text{where} \quad B \;=\; \begin{pmatrix} \boldsymbol{e}_h{}^T \\ \\ \boldsymbol{p}^T \end{pmatrix} \;=\; \begin{pmatrix} 0 & \ldots & 1 & \ldots & 0 \\ \\ p_1 & \ldots & p_h & \ldots & p_{\lfloor H \rfloor} \end{pmatrix}\;.$$

If we have $\boldsymbol{p} = \mathbf{0}$, then (3.6) obviously reduces to $-q_k\,\mathrm{E}(\,\phi_h\,) \;=\; -q_k\,\mu_h$.

If we have $\boldsymbol{p}^T = (0, \ldots, p_h, \ldots, 0)$, $p_h \neq 0$, then we only need to consider the random variable ϕ_h. It has a normal distribution with expectation μ_h and variance c_{hh}. Let $f(x)$ denote the density function of this distribution. Then (3.6) reduces to

$$-q_k \int\limits_{-\infty}^{b_k/p_h} x\; f(x)\; dx \;=\; q_k\;\frac{\sqrt{c_{hh}}}{\sqrt{2\pi}}\; e^{-\frac{1}{2c_{hh}}\,(b_k/p_h - \mu_h)^2} \;-\; q_k\;\mu_h\;\Phi\left(\frac{b_k/p_h \;-\; \mu_h}{\sqrt{c_{hh}}}\right)\;.$$

If, on the other hand, the matrix B has rank 2, then the random vector $\boldsymbol{\psi}$ has a non-degenerate 2-variate normal distribution. The expectation vector and the covariance matrix

of $\boldsymbol{\psi}$ are $B\boldsymbol{\mu}$ and BCB^T, respectively. For the expectation vector, let us introduce the notation

$$\begin{pmatrix} \nu_1 \\ \\ \nu_2 \end{pmatrix} \;=\; B\boldsymbol{\mu} \;=\; \begin{pmatrix} \boldsymbol{e}_h{}^T\boldsymbol{\mu} \\ \\ \boldsymbol{p}^T\boldsymbol{\mu} \end{pmatrix} \;=\; \begin{pmatrix} \mu_h \\ \\ \mu \end{pmatrix} \;.$$

For the covariance matrix, let us inroduce the notation

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \;=\; BCB^T \;=\; \begin{pmatrix} \boldsymbol{e}_h{}^TC\boldsymbol{e}_h & \boldsymbol{e}_h{}^TC\boldsymbol{p} \\ \\ \boldsymbol{p}^TC\boldsymbol{e}_h & \boldsymbol{p}^TC\boldsymbol{p} \end{pmatrix} \;=\; \begin{pmatrix} c_{h,h} & \boldsymbol{e}_h{}^TC\boldsymbol{p} \\ \\ \boldsymbol{p}^TC\boldsymbol{e}_h & \sigma^2 \end{pmatrix} \;.$$

Let us designate by $f(x,y)$ the density function of the random vector $\boldsymbol{\psi}$, and write (3.6) in the following form:

$$-q_k \int_{-\infty}^{b_k} \int_{-\infty}^{\infty} x\; f(x,y)\; dx\; dy \;\;. \tag{3.7}$$

Let further $f(\,x \mid y\,)$ designate the conditional density function of ϕ_h given $\boldsymbol{p}^T\boldsymbol{\phi} \,=\, y$; and $g(y)$ the density function of $\boldsymbol{p}^T\boldsymbol{\phi}$. We have the equality

$$f(\,x,\,y\,) \;\;=\;\; f(\,x \mid y\,)\, g(\,y\,) \;\;.$$

The random variable ϕ_h given that $\boldsymbol{p}^T\boldsymbol{\phi} = y$ is known to have a normal distribution with expectation

$$\nu_1 \;+\; \rho\,\frac{\sigma_1}{\sigma_2}\,(\,y - \nu_2\,) \;\;,$$

and variation $\sigma_1^2\,(1-\rho^2\,)$. (See Section 7.3 of Wilks (1962), or for the general case, Theorem 5 of Appendix in Prékopa (1995).)

Substituting the above expressions into (3.7), we get

$$-q_k \int_{-\infty}^{b_k} \int_{-\infty}^{+\infty} x\; f(\,x \mid y\,)\; dx\;\; g(\,y\,)\; dy \;\;=\;\; -q_k \int_{-\infty}^{b_k} \Big(\,\nu_1 \;+\; \rho\,\tfrac{\sigma_1}{\sigma_2}\,(\,y - \nu_2\,)\Big)\;\; g(\,y\,)\; dy$$

$$=\;\; -q_k\,\nu_1 \int_{-\infty}^{b_k} g(y)\; dy \;\;+\;\; q_k\,\tfrac{\rho\,\sigma_1}{\sqrt{2\pi}} \int_{-\infty}^{b_k} -\tfrac{1}{\sigma_2^2}\,(y-\nu_2)\; e^{\,-\frac{1}{2\sigma_2^2}\,(y-\nu_2)^2}\; dy$$

$$=\;\; -q_k\,\nu_1\,\Phi\Big(\tfrac{b_k-\nu_2}{\sigma_2}\Big) \;\;+\;\; q_k\,\tfrac{\rho\,\sigma_1}{\sqrt{2\pi}}\; e^{\,-\frac{1}{2\sigma_2^2}\,(b_k-\nu_2)^2} \;\;.$$

Thus we can compute the gradient of the objective function.

3.4 Outline of a Solution Method

Our aim is to solve problem (3.1). We have seen in Section 3.2 that the objective function is strictly convex and continuously differentiable. On the other hand, the convex polyhedron P is not known explicitly.

We use a feasible direction method, Zoutendijk's procedure P2 (1960). All constraints are linear, hence procedure P2 can be simplified. In the following description, points of the polyhedron P will be considered vectors rather than matrices and will be denoted by $\boldsymbol{a}$ rather than by A. The gradient of the objective function at $\boldsymbol{a} \in \mathbb{R}^{m\lfloor H \rfloor}$ will be denoted by $\boldsymbol{g}(\boldsymbol{a})$. The solution algorithm consists of the following steps:

{0} Find an initial vector $\boldsymbol{a} \in P$.

{1} Find a feasible direction $\boldsymbol{d} \in \mathbb{R}^{m\lfloor H \rfloor}$

which minimizes the directional derivative $\Big[\, \boldsymbol{g}(\boldsymbol{a}) \,\Big]^T \boldsymbol{d}$

subject to $\boldsymbol{a} + \boldsymbol{d} \in P$, $\| \boldsymbol{d} \|_{\max} \leq 1$.

If the optimal directional derivative is small in absolute value, then the point $\boldsymbol{a}$ is near-optimal for the original problem.

{2} Determine the step length, i.e., find $\lambda \in \mathbb{R}$

which minimizes the objective value $\Delta(\boldsymbol{a} + \lambda \boldsymbol{d})$

subject to $\boldsymbol{a} + \lambda \boldsymbol{d} \in P$, $\lambda \geq 0$.

Set $\boldsymbol{a} := \boldsymbol{a} + \lambda \boldsymbol{d}$.

Steps {1} and {2} are iterated until near-optimal solution is found. Near-optimality is tested by using some pre-defined tolerance.

Zoutendijk proves that under certain conditions procedure P2 is convergent. These conditions are met in the present case, because all the contstraints are linear, the feasible region is compact, the objective function is strictly convex and continuously differentiable. Convergence holds in the following sense: suppose the stopping tolerance is set to zero, and the procedure produces the sequence of feasible points $\Big(\boldsymbol{a}^\ell,\ \ell = 1, 2, \ldots\Big)$. Then the sequence of the corresponding objective values $\Big(\Delta\Big(\boldsymbol{a}^\ell\Big)\Big)$ converges to the optimal value. Since the present objective function is strictly convex, and the feasible region compact, it follows that the sequence $\Big(\boldsymbol{a}^\ell\Big)$ is also convergent.

Now, let us turn our attention to the linear programming problem to be solved in step {1} of the procedure. Constraints of that problem involve the polyhedron $P = P_1 \times \cdots \times P_{\lfloor H \rfloor}$.

Since we use the maximum norm for the normalization of the direction $\boldsymbol{d}$, the linear programming problem {1} breaks up into small subproblems in accordance with the direct product. Consider the projection of the direct product into one of its components P_h. Extend that projection to the whole space and denote it by $\pi_h : \mathbb{R}^{m\lfloor H \rfloor} \longrightarrow \mathbb{R}^m$. Introduce the notations

$$\boldsymbol{a}_h = \pi_h(\boldsymbol{a}), \quad \boldsymbol{g}_h(\boldsymbol{a}) = \pi_h\Big(\, \boldsymbol{g}(\boldsymbol{a}) \,\Big) .$$

We solve $\lfloor H \rfloor$ subproblems, of which the hth one is the following:

$\{1.h\}$ Find $\boldsymbol{d}_h \in \mathbb{R}^m$

which minimizes the directional derivative $\left[\, \boldsymbol{g}_h(\boldsymbol{a}) \,\right]^T \boldsymbol{d}_h$

subject to $\boldsymbol{a}_h + \boldsymbol{d}_h \in P_h$, $\| \boldsymbol{d}_h \|_{\max} \leq 1$.

The optimal solutions $\boldsymbol{d}_h$ of the subproblems $\{1.h\}$ $(h = 1, \ldots, \lfloor H \rfloor)$ jointly give an optimal solution $\boldsymbol{d}$ for the linear programming problem in step $\{1\}$.

Since the convex polyhedron P_h is the convex hull of the cutting patterns for length h, the number of its facets may be excessively large. The subproblems can be solved by the method of Gilmore and Gomory (1961). Below we briefly describe the method as applied to problem $\{1.h\}$.

$\{1.h.0\}$ At the beginning, it is enough to know only a few vertices of the polyhedron P_h . Denote the convex hull of the known vertices by Q_h .

$\{1.h.1\}$ Solve a simplified version of the problem $\{1.h\}$. Minimize the objective function over the convex hull Q_h , instead of the convex polyhedron P_h . The modified constraints are $\boldsymbol{a}_h + \boldsymbol{d}_h \in Q_h$, $\| \boldsymbol{d}_h \|_{\max} \leq 1$.

$\{1.h.2\}$ Find a new vertex $\boldsymbol{p}_h$ of P_h by which to extend the convex hull Q_h to cause further decrease in the objective function of the problem $\{1.h.1\}$. I.e., the optimal objective value is smaller on $\mathrm{Conv}(\boldsymbol{p}_h, Q_h)$ then on Q_h .

If no such vertex exists, then optimal solution of problem $\{1.h\}$ has been found. If an improving vertex is found, then let $Q_h := \mathrm{Conv}(\boldsymbol{p}_h, Q_h)$.

Steps $\{1.h.1\}$ and $\{1.h.2\}$ have to be iterated until all the necessary vertices are discovered. Finding a new vertex of P_h means generating a new cutting pattern for length h. Improving patterns can be discovered by the use of the dual variables of the linear programming problem $\{1.h.1\}$, as follows. Suppose that after n iterations of steps $\{1.h.1\}$ and $\{1.h.2\}$, we have discovered the vertices $\boldsymbol{p}_h^1, \ldots, \boldsymbol{p}_h^n$ of the convex polyhedron P_h . Hence we have $Q_h = \mathrm{Conv}(\boldsymbol{p}_h^1, \ldots, \boldsymbol{p}_h^n)$. Thus the constraint $\boldsymbol{a}_h + \boldsymbol{d}_h \in Q_h$ in problem $\{1.h.1\}$ means that there exist non-negative real numbers $\lambda_1, \ldots, \lambda_n$ such that

$$\begin{aligned} \sum_{i=1}^{n} \lambda_i & = 1 \\ \sum_{i=1}^{n} \lambda_i \boldsymbol{p}_h^i - \boldsymbol{d}_h & = \boldsymbol{a}_h \ . \end{aligned}$$

Let $y_0 \in \mathbb{R}$, $\boldsymbol{y} \in \mathbb{R}^m$ be the dual variables corresponding to the above $1 + m$ equalities. If we found an optimal solution of problem $\{1.h.1\}$, then we have

$$y_0 + \boldsymbol{y}^T \boldsymbol{p}_h^i \geq 0 \qquad (i = 1, \ldots, n).$$

Problem $\{1.h.2\}$ can be formulated in terms of the dual variables as

Find a cutting pattern $\boldsymbol{p}_h \in P_h$ which minimizes $\boldsymbol{y}^T \boldsymbol{p}_h$.

This can be done by solving a knapsack problem, the constraint of which is inequality (2.2) that characterizes the cutting patterns. If the optimal objective value of the above problem is not smaller then $-y_0$, then no improving pattern exists. The column generation method of Gilmore and Gomory is widely used, and considered efficient.

Remark. The convex polyhedron P is bounded. Hence the method remains convergent, if we simplify problem $\{1.h\}$ by removing the normalization constraint $\| \boldsymbol{d}_h \|_{\max} \leq 1$. In this case, problem $\{1.h\}$ reduces to the minimization of a linear function over the convex polyhedron

$$P_h - \boldsymbol{a}_h \;=\; \left\{ \boldsymbol{x} - \boldsymbol{a}_h \;\middle|\; \boldsymbol{x} \in P_h \right\} \;.$$

An optimal solution corresponds to a vertex of the above polyhedron, which, in turn, corresponds to a cutting pattern for length h . Hence problem $\{1.h\}$ reduces to a single knapsack problem. □

Finally, as regards the problem of step length determination $\{2\}$, any method can be used. The only irregularity is in the determination of the feasible interval. This requires the solution of a parametric linear programming problem over the polyhedron P , and can be treated like problem $\{1\}$. If, furthermore, we remove the normalization constraint, then the length of the feasible interval will always be 1. I.e., the step length determination problem reduces to the minimization of $\Delta(\boldsymbol{a} + \lambda \boldsymbol{d})$ subject to $0 \leq \lambda \leq 1$.

Chapter 4

Reliability Considerations for the Static Model

If we select the cutting rule that minimizes the expected loss, big losses may occur with a probability higher than tolerable. We may have to sacrifice some of the expected profit for a higher level of reliability.

Murr and Prékopa (1996) use probabilistic constraints in a joint constraint form to ensure that all demands are satisfied with a high probability. In the present model, reliability is ensured by imposing some upper bound on the second moment of the loss. Formally, denote the square of loss by

$$\delta^{(2)}(A, \boldsymbol{\phi}) \;=\; \Big(\delta(A, \boldsymbol{\phi})\Big)^2 \;=\; \Big(\boldsymbol{q^T}\,[\,\boldsymbol{b} - A\boldsymbol{\phi}\,]_+\Big)^2 \;,$$

and the second moment of the loss by

$$\Delta^{(2)}(A) \;=\; \mathrm{E}\Big(\delta^{(2)}(A, \boldsymbol{\phi})\Big) \;.$$

Imposing the upper bound $U^{(2)} \in \mathbb{R}$ on the second moment, the problem is

$$\begin{aligned} A &\in P \\ \Delta^{(2)}(A) &\leq U^{(2)} \\ \min \quad \Delta(A) \quad &. \end{aligned} \tag{4.1}$$

(In another approach, we might impose some upper bound on the expected loss, and minimize the second moment of the loss.)

The upper bound $U^{(2)} \in \mathbb{R}$ is a parameter that should be set according to experience. For

$$U^{(2)} \;>\; \min_{A \in P} \Delta^{(2)}(A) \;,$$

the feasible region is not empty.

We show that the function $\Delta^{(2)}$ is strictly convex and continuously differentiable. Due to the strict convexity of the objective function Δ, and the convexity of the constraint function $\Delta^{(2)}$; the optimal solution $A^\star$ of the problem (4.1) is efficient in the sense of Markowitz (1952). That is, there does not exist any cutting rule A for which we have $\Delta(A) = \Delta(A^\star)$, $\Delta^{(2)}(A) < \Delta^{(2)}(A^\star)$ or $\Delta^{(2)}(A) = \Delta^{(2)}(A^\star)$, $\Delta(A) < \Delta(A^\star)$.

In Section 4.2, we propose a solution method. Zoutendijk's feasible direction method P2 can be used, and the feasible region can be explored by the method of Gilmore and Gomory; just like in the case of the original problem.

4.1 Properties of the Function $\Delta^{(2)}$

We show that the function is strictly convex and continuously differentiable.

For a fixed yield $\boldsymbol{\phi}$, the function $\delta^{(2)}(\,.\,, \boldsymbol{\phi})$ is convex and almost everywhere differentiable in the first variable $A \in \mathbb{R}^{m \times \lfloor H \rfloor}$. (Namely, it is differentiable if the vector $[\,\boldsymbol{b} - A\boldsymbol{\phi}\,]_+$ falls into the interior of any of the orthants of the space $\mathbb{R}^m$.)

For any $A \in \mathbb{R}^{m \times \lfloor H \rfloor}$, pick a subgradient of the above function, and denote it by $\boldsymbol{\gamma}^{(2)}(\,A,\, \boldsymbol{\phi}\,)$. That subgradient is also gradient for almost every A.

Suppose that rule A and yield $\boldsymbol{\phi}$ are such that the function $\delta^{(2)}(\,.\,, \boldsymbol{\phi}\,)$ is differentiable in A. We will express the gradient as a function of A and $\boldsymbol{\phi}$. For that end, let us split the function δ as we did before:

$$\delta\,(\,A,\, \boldsymbol{\phi}\,) \;=\; \sum_{k=1}^{m} \;\delta_k\left(\boldsymbol{a^k},\, \boldsymbol{\phi}\right) \quad ,$$

where $\boldsymbol{a}^k$ is the kth row of the matrix A, and $\delta_k\left(\boldsymbol{a^k},\, \boldsymbol{\phi}\right) = q_k \left[\, b_k - \boldsymbol{a^k\phi}\,\right]_+$.

Hence the function $\delta^{(2)}$ splits as follows:

$$\begin{aligned}\delta^{(2)}\,(\,A,\, \boldsymbol{\phi}\,) \;&=\; \sum_{k=1}^{m} \left(\,\delta_k\left(\boldsymbol{a^k},\, \boldsymbol{\phi}\right)\right)^2 \;+\; \sum_{\substack{k,\,\ell\,=\,1\\ k\,\neq\,\ell}}^{m} \delta_k\left(\boldsymbol{a^k},\, \boldsymbol{\phi}\right)\,\delta_\ell\left(\boldsymbol{a^\ell},\, \boldsymbol{\phi}\right) \\ &=\; \sum_{k=1}^{m} \left(\,q_k \left[\, b_k - \boldsymbol{a}^k\boldsymbol{\phi}\,\right]_+\right)^2 \;+\; \sum_{\substack{k,\,\ell\,=\,1\\ k\,\neq\,\ell}}^{m} q_k\, q_\ell \left[\, b_k - \boldsymbol{a}^k\boldsymbol{\phi}\,\right]_+ \left[\, b_\ell - \boldsymbol{a}^\ell\boldsymbol{\phi}\,\right]_+ \quad .\end{aligned} \tag{4.2}$$

Let us express the gradient of the function $\delta^{(2)}(\,.\,, \boldsymbol{\phi}\,)$ componentwise : Fix the matrix A with the exception of the kth row. (That is, fix $\boldsymbol{a}^\ell$ $\ell = 1, \ldots, m;\ \ell \neq k$.) The variable will be $\boldsymbol{a}^k$, the kth row of the matrix.

The gradient of the term $\left(\delta_k \left(\boldsymbol{a^k}, \boldsymbol{\phi} \right) \right)^2$ is the following

$$\begin{cases} \mathbf{0} \in \mathbb{R}^{\lfloor H \rfloor} \ , & \text{if } b_k < \boldsymbol{a^k \phi} \ ; \\ -2q_k^2 \left(b_k - \boldsymbol{a}^k \boldsymbol{\phi} \right) \boldsymbol{\phi} \ , & \text{if } b_k > \boldsymbol{a^k \phi} \ . \end{cases} \tag{4.3}$$

Similarly, the gradient of the term $\delta_k \left(\boldsymbol{a^k}, \boldsymbol{\phi} \right) \delta_\ell \left(\boldsymbol{a^\ell}, \boldsymbol{\phi} \right)$ is

$$\begin{cases} \mathbf{0} \in \mathbb{R}^{\lfloor H \rfloor} \ , & \text{if } \ b_k < \boldsymbol{a}^k \boldsymbol{\phi} \ \text{ or } \ b_\ell \leq \boldsymbol{a}^\ell \boldsymbol{\phi} \ ; \\ -q_k \, q_\ell \left(b_\ell - \boldsymbol{a}^\ell \boldsymbol{\phi} \right) \boldsymbol{\phi} \ , & \text{if } \ b_k > \boldsymbol{a}^k \boldsymbol{\phi} \ \text{ and } \ b_\ell > \boldsymbol{a}^\ell \boldsymbol{\phi} \ . \end{cases} \tag{4.4}$$

We may assume $b_k, b_\ell > 0$. Hence for any vector $\boldsymbol{a}^k$, formulae (4.3) and (4.4) determine the respective gradients for almost every $\boldsymbol{\phi}$.

It is easy to verify that the expectations of the gradient vectors (4.3) and (4.4) exist, and are continuous functions of the vector $\boldsymbol{a}^k$. It follows that for any rule A, the expectation of the subgradient $\boldsymbol{\gamma}^{(2)}(A, \boldsymbol{\phi})$ exists, and is a continuous function of A.

Like in the case of the first moment (the Δ function), follows that the function $\Delta^{(2)}$ is convex and differentiable. Namely, at A, the gradient is the expectation of $\boldsymbol{\gamma}^{(2)}(A, \boldsymbol{\phi})$.

Remark. The function $\Delta^{(2)}$ is even strictly convex. The proof is the same as in the case of the first moment. □

Both the function value and the gradient can be computed by the integration of the one- and two-dimensional normal probability density functions. The computation is similar to that of Δ.

4.2 Solution Method for the Modified Problem

Our aim is to solve problem (4.1). Both the objective function and the constraint function $\Delta^{(2)}$ are strictly convex and continuously differentiable. On the other hand, the convex polyhedron P is not known explicitly.

We use Zoutendijk's algorithm P2 (1960). In the following description, points of the polyhedron P will be considered vectors rather than matrices and will be denoted by $\boldsymbol{a}$ rather than by A.

The gradient of the objective function Δ at $\boldsymbol{a} \in \mathbb{R}^{m\lfloor H \rfloor}$ will be denoted by $\boldsymbol{g}(\boldsymbol{a})$. SImilarly, the gradient of the function $\Delta^{(2)}$ will be denoted by $\boldsymbol{g}^{(2)}(\boldsymbol{a})$. The solution algorithm consists of the following steps:

$\{\overline{0}\}$ Find an initial vector $\boldsymbol{a} \in P$, $\Delta^{(2)}(\boldsymbol{a}) \leq U^{(2)}$.

$\{\overline{1}\}$ Find a feasible direction $\boldsymbol{d} \in \mathbb{R}^{m\lfloor H \rfloor}$ such that $\boldsymbol{a} + \boldsymbol{d} \in P$, $\| \boldsymbol{d} \|_{\max} \leq 1$, which maximizes $x \in \mathbb{R}$ (the measure of the usability of the direction)

subject to the following inequalities for the linear approximations of the functions $\Delta^{(2)}$ and Δ :

$$\begin{aligned} \Delta^{(2)}(\boldsymbol{a}) \; + \; \left[\boldsymbol{g}^{(2)}(\boldsymbol{a}) \right]^T \boldsymbol{d} \;\; &\leq \;\; U^{(2)} \;\; - x \\ \left[\boldsymbol{g}(\boldsymbol{a}) \right]^T \boldsymbol{d} \;\; &\leq \qquad\quad - x \;\; . \end{aligned} \tag{4.5}$$

If the optimal value of x is small, then the point $\boldsymbol{a}$ is near-optimal for the original problem.

$\{\overline{2}\}$ Determine step length, i.e., find $\lambda \in \mathbb{R}$

which minimizes objective value $\Delta(\boldsymbol{a} + \lambda \boldsymbol{d})$,

subject to $\boldsymbol{a} + \lambda \boldsymbol{d} \in P$, $\Delta^{(2)}(\boldsymbol{a} + \lambda \boldsymbol{d}) \leq U^{(2)}$, $\lambda \geq 0$.

Set $\boldsymbol{a} := \boldsymbol{a} + \lambda \boldsymbol{d}$.

Steps $\{\overline{1}\}$ and $\{\overline{2}\}$ are iterated until near-optimal solution is found. Near-optimality is tested by using some pre-defined tolerance.

Zoutendijk's convergence conditions are the following:

- For each nonlinear constraint, there must exist a feasible point in which the constraint holds with strict inequality.
- The set of those feasible points, which has optimal objective value, must be bounded.

In the present case the above conditions are clearly satisfied, because the objective function is strictly convex, and the upper bound $U^{(2)}$ is strictly larger than the minimum of the second moment over the polyhedron of cutting rules. Hence the method is convergent.

Remark. Since the polyhedron P is bounded, the method remains convergent, if we discard the normalization constraint $\|\boldsymbol{d}\|_{\max} \leq 1$ in problem $\{\overline{1}\}$. We will do so, in order to simplify the following discussion. □

The problem of step $\{\overline{0}\}$ can be solved by the method proposed for the original simple recourse problem.

The direction finding problem of step $\{\overline{1}\}$ has a special block structure. The two constraints of (4.5) hold the subproblems together.

We formulate the problem $\{\overline{1}\}$ with an emphasis on the block structure. Denote by π_h the projection $\mathbb{R}^{m\lfloor H \rfloor} \longrightarrow \mathbb{R}^m$ that projects the direct product $P_1 \times \cdots \times P_{\lfloor H \rfloor}$ into P_h . Let us introduce the variable vectors

$$\boldsymbol{p}_h = \pi_h (\boldsymbol{a} + \boldsymbol{d}) \qquad (h = 1, \ldots, \lfloor H \rfloor)$$

instead of the vector $\boldsymbol{d}$. With this transformation, the direction finding problem $\{\overline{1}\}$ can be written as follows

$$\text{find } \boldsymbol{p}_h \in P_h\,,\ h=1,\ldots,\lfloor H\rfloor \quad \text{which maximizes } x\in\mathbb{R}$$

$$\text{subject to } \sum_{h=1}^{\lfloor H\rfloor}\Big[\pi_h\Big(\boldsymbol{g}^{(2)}(\boldsymbol{a})\Big)\Big]^T\boldsymbol{p}_h \;+\; x \;\le\; -\Delta^{(2)}(\boldsymbol{a}) \;+\; U^{(2)} \;+\; \Big[\boldsymbol{g}^{(2)}(\boldsymbol{a})\Big]^T\boldsymbol{a}$$

$$\sum_{h=1}^{\lfloor H\rfloor}\Big[\pi_h\Big(\boldsymbol{g}(\boldsymbol{a})\Big)\Big]^T\boldsymbol{p}_h \;+\; x \;\le\; \Big[\boldsymbol{g}(\boldsymbol{a})\Big]^T\boldsymbol{a}\ .$$

The problem can be solved effectively using the decomposition method of Dantzig and Wolfe (1961). (The decomposition method of Dantzig and Wolfe, and the column generation method of Gilmore and Gomory overlap in the case of this particular problem.) The method is the following

$\{\overline{1.0}\}$ At the beginning, it is enough to know only a few vertices of each polyhedron P_h. Denote the convex hulls of the known vertices by Q_h, $(h=1,\ldots,\lfloor H\rfloor)$.

$\{\overline{1.1}\}$ Solve a simplified version of the direction finding problem, imposing the constraint $\boldsymbol{p}_h\in Q_h$ instead of $\boldsymbol{p}_h\in P_h$ $(h=1,\ldots,\lfloor H\rfloor)$.

$\{\overline{1.2}\}$ Find a new vertex $\boldsymbol{p}_h$ of P_h by which to extend the convex hull Q_h to cause further decrease in the objective function of the problem $\{\overline{1.1}\}$. I.e., the optimal objective value is smaller on $\mathrm{Conv}(\boldsymbol{p}_h,\,Q_h)$ then on Q_h.

If no such vertex exists, then optimal solution of problem $\{\overline{1.1}\}$ has been found. If an improving vertex is found, then let $Q_h := \mathrm{Conv}(\boldsymbol{p}_h,\,Q_h)$.

Steps $\{\overline{1.0}\}$ and $\{\overline{1.1}\}$ are iterated until all the necessary vertices are discovered.

Now we describe problem $\{\overline{1.1}\}$ in detail. Suppose that for length $1\le h\le\lfloor H\rfloor$, we already discovered the vertices (that is, cutting patterns)

$$\boldsymbol{p}_h^1\,,\ \ldots,\ \boldsymbol{p}_h^{s_h} \;\in\; P_h\ .$$

The constraint $\boldsymbol{p}_h\in Q_h$ for the variable vector $\boldsymbol{p}_h\in\mathbb{R}^m$ is equivalent to the existence of a non-negative vector $\boldsymbol{\lambda}_h = (\lambda_h^1,\ldots,\lambda_h^{s_h})^T$, such that

$$\boldsymbol{p}_h \;=\; \sum_{\ell=1}^{s_h}\lambda_h^\ell\,\boldsymbol{p}_h^\ell\ ,\qquad \lambda_h^1+\ldots+\lambda_h^{s_h} \;=\; 1\ .$$

Let us introduce the notations

$$u_h^\ell \;=\; \Big[\pi_h\Big(\boldsymbol{g}^{(2)}(\boldsymbol{a})\Big)\Big]^T\boldsymbol{p}_h\,,\quad v_h^\ell \;=\; \Big[\pi_h\Big(\boldsymbol{g}(\boldsymbol{a})\Big)\Big]^T\boldsymbol{p}_h \qquad (\ell=1,\ldots,s_h)\ ,$$

and

$$\boldsymbol{u}_h \;=\; (u_h^1,\ldots,u_h^{s_h})^T\,,\quad \boldsymbol{v}_h \;=\; (v_h^1,\ldots,v_h^{s_h})^T\ .$$

With the above notations, the simplified direction finding problem $\{\overline{1.1}\}$ can be written as follows

$$\text{find } \boldsymbol{\lambda}_h \in \mathbb{R}^{s_h} \text{ such that } \boldsymbol{\lambda}_h \geq \mathbf{0},\ \mathbf{1}^T\boldsymbol{\lambda}_h = 1, \qquad h = 1, \ldots, \lfloor H \rfloor$$

$$\text{which maximizes } x \in \mathbb{R}$$

$$\text{subject to } \sum_{h=1}^{\lfloor H \rfloor} \boldsymbol{u}_h{}^T \boldsymbol{\lambda}_h + x \leq -\Delta^{(2)}(\boldsymbol{a}) + U^{(2)} + \left[\boldsymbol{g}^{(2)}(\boldsymbol{a})\right]^T \boldsymbol{a}$$

$$\sum_{h=1}^{\lfloor H \rfloor} \boldsymbol{v}_h{}^T \boldsymbol{\lambda}_h + x \leq \left[\boldsymbol{g}(\boldsymbol{a})\right]^T \boldsymbol{a} \ .$$

The above problem will be the master problem in the Dantzig-Wolfe decomposition. We have $\lfloor H \rfloor$ subproblems that generate new improving cutting patterns for the different lengths $h = 1, \ldots, \lfloor H \rfloor$. Each subproblem is a knapsack problem, the constraint of which is the inequality which characterizes cutting patterns. The objective function of the knapsack problem is composed from the dual variables of the master problem.

Chapter 5

A Dynamic Model

In this chapter we formulate a two-period model. In practice, rolling horizon scheme can be used: First we formulate the problem for periods 1 and 2 but accept as final the values of the decision variables for period 1 only. Then after period 1, we formulate the problem for periods 2 and 3 etc.

Our two-stage model is a special case of the two-stage stochastic programming problem. See Chapter 12 of Prékopa (1995). The decision/observation scheme in the present case is the following:

– Decision on the cutting rule to be used in period 1.

– Observation of the first-period random yield, i.e., the set of semi-finished products produced during period 1. (The cutting rule and the yield together determine the numbers of different products produced during period 1.)

– Decision on selling or storing products in stock at the end of period 1, and on the cutting rule to be used in period 2.

– Observation of the second-period yield.

– Decision on selling products in stock at the end of period 2. This final decision is simple: in each product, as many pieces of fiber must be sold as the market can take. (We ignore the possibility of storing products beyond period 2.)

The decision problem which must be solved at the beginning of period 2 is called the *second-stage problem*, and the decision problem which must be solved at the beginning of period 1 is called the *first-stage problem*.

Under reasonable assumptions, the second-stage problem reduces to finding a second-period cutting rule that minimizes the expectation of the second-period loss. This is equivalent to the one-period problem formulated and solved in Chapter 3.

The above model has a special feature: the first-period cutting rule and yield affect the second-stage problem only through the inventory of the products in stock at the beginning of the second period. The inventory can be described by a relatively small amount of data.

We consider the optimal objective value of the second-stage problem (i.e., the minimum of the expectation of the second-period loss) as a function of the inventory. This function will be referred to as the *inventory function*. The inventory function is smooth (continuously differentiable). Given an inventory vector, both the function value and the gradient can be computed.

The objective function of the first-stage problem is convex. Given a first-period cutting rule, the objective function can be evaluated by computing the expectation of the inventory function value. (Expectation is considered with respect to a random distribution which depends on the first-period cutting rule.)

We propose to estimate the objective value by simulation. To enable simulation, the inventory function is approximated by piecewise linear, convex functions which are easy to evaluate. Both lower and upper approximating functions are constructed. Hence given a first-period cutting rule, both lower and upper estimators can be found for the objective value. (The expectations of piecewise linear functions must be considered with respect to a random distribution which depends on the cutting rule.) We work out the details for lower approximation. Upper estimators will be used only to measure accuracy.

Consider a piecewise linear, convex, lower approximate for the inventory function. In the way sketched above, this defines a lower approximate for the objective function. The approximate objective function is convex, and at a given point, a subgradient can be estimated by simulation.

Consider the function value and subgradient of a lower approximate objective function at a given point (i.e., first-period cutting rule). These data will be used to construct a linear function which underestimates the original objective function. The upper cover of such linear functions is a convex, piecewise linear function which underestimates the objective function. The more linear functions are constructed, the better estimator the upper cover is.

The minimization problem is similar to the static problem described in Chapter 3. In the case of the dynamic problem, however, finding a good estimate of the objective value and subgradient requires a relatively big effort. Hence the solution method proposed for the static problem is not effective in the present case, because a feasible direction method may take a large number of steps to approach optimal solution with prescribed precision. A suitable solution method is presented in Chapter 8.

The optimization problem is formulated in Section 5.1. In Section 5.2 the mathematical properties of the objective function are derived. In Section 5.3 a lower approximation is presented for the objective function. Accuracy of the lower approximation will be measured by computing upper bounds for the objective value at a given point. A method for computing upper bounds is presented in Section 5.4.

5.1 Formulation of the Problem

Let $b_k{}'$ designate the first-period demand for products of length h_k $(k = 1, \ldots, m)$. Let us form the vector $\boldsymbol{b}' = (b_1{}', \ldots, b_m{}')^T$. Similarly, let $\boldsymbol{b} = (b_1, \ldots, b_m)^T$ designate the

second-period demand in different products.

We assume that at the start of the first period, there are no products in store. Let the m-vector $\boldsymbol{v}$ designate the numbers of the different products in store at the beginning of the second period. This vector will be referred to as the (second-period) *inventory*. To clarify notations, let us review the decision/observation scheme :

– Cutting rule A' was selected for period 1.

– Yield $\boldsymbol{\phi}'$ was produced during period 1. (The vector $A'\boldsymbol{\phi}'$ represents the numbers of the different products cut from this yield.)

– Between periods 1 and 2, decision was made on the numbers of products to be stored for future sale. (The inventory $\boldsymbol{v}$ must satisfy the inequality $[\, A'\boldsymbol{\phi}' \,]_+ \geq \boldsymbol{v} \geq \mathbf{0}$. Under certain reasonable assumptions, the optimal decision turns out to be the following: in each product, as many pieces of fiber must be sold as the market can take. Surplus products must be stored for future sale.)
At the same time, cutting rule A was selected for period 2.

– Yield $\boldsymbol{\phi}$ was produced during period 2. (The vector $A\boldsymbol{\phi}$ represents the numbers of the different products cut from this yield.)

At the end of the second period, the unsatisfied demands for the different products are the components of the vector

$$[\, \boldsymbol{b} - \boldsymbol{v} - A\boldsymbol{\phi} \,]_+ \quad .$$

Let

$$\boldsymbol{q}^T [\, \boldsymbol{b} - \boldsymbol{v} - A\boldsymbol{\phi} \,]_+$$

be the loss (unrealized profit) due to unsatisfied demand, where $\boldsymbol{q} \geq \mathbf{0}$ is a given second-period cost vector.

The second-stage problem is the following. Choose the second-period inventory $\boldsymbol{v}$ and the second-period cutting rule A which minimizes the sum of the first-period loss and the expectation of the second-period loss, i.e.,

$$\boldsymbol{q}'^T [\, \boldsymbol{b}' - A'\boldsymbol{\phi}' + \boldsymbol{v} \,]_+ + \underbrace{\mathrm{E}\Big(\boldsymbol{q}^T [\, \boldsymbol{b} - A\boldsymbol{\phi} - \boldsymbol{v} \,]_+ \Big)}_{\Delta(\boldsymbol{v}, A)} , \tag{5.1}$$

where $\boldsymbol{q}' \geq \mathbf{0}$ is a given first-period cost vector, and the expectation is considered with respect to the random vector $\boldsymbol{\phi}$. Let $\Delta(\boldsymbol{v}, A)$ represent this expectation, as indicated.

We assume $\boldsymbol{q}' \geq \boldsymbol{q}$. Under this assumption, only surplus products are to be stored for future sale (i.e., products that can not be sold during the first period). The second-period inventory vector which minimizes (5.1) is

$$\boldsymbol{v} = [\, A'\boldsymbol{\phi}' - \boldsymbol{b}' \,]_+ \ .$$

Let us substitute the optimal second-period inventory into (5.1). The second-stage problem reduces to finding the second-period cutting rule A which minimizes

$$\boldsymbol{q}'^T [\, A'\boldsymbol{\phi}' - \boldsymbol{b}' \,]_- + \Delta\Big([\, A'\boldsymbol{\phi}' - \boldsymbol{b}' \,]_+ , A \Big) . \tag{5.2}$$

The first term of the above sum does not depend on A. Minimization of the second term over the convex polyhedron of the cutting rules means the solution of a one-period problem. We treated such problems in and proposed a solution method for them. The first-period cutting rule and yield affect this one-period problem only through the inventory vector. Let $r(\boldsymbol{v})$ designate the optimal objective value of the one-period problem, as a function of the inventory :

$$r(\boldsymbol{v}) = \min_{A \in P} \Delta(\boldsymbol{v}, A) .$$

The function r will be referred to as the *inventory function.* With this notation, the optimal value of (5.2) can be written as

$$\theta(A', \boldsymbol{\phi}') = \boldsymbol{q}'^T [\, A'\boldsymbol{\phi}' - \boldsymbol{b}' \,]_- + r([\, A'\boldsymbol{\phi}' - \boldsymbol{b}' \,]_+) .$$

Now we can formulate the problem to be solved at the start of the first period, called the first-stage problem. We must select the cutting rule A' which minimizes the expectation

$$\Theta(A') = \mathrm{E}\Big(\theta(A', \boldsymbol{\phi}') \Big) .$$

Expectation is considered with respect to the random vector $\boldsymbol{\phi}'$. By intuition, the expectation exists and is finite. A formal proof is given in Section 5.2.

5.2 Properties of the Objective Function

Our aim is to show that the objective function is convex, and to examine differentiability.

Let us extend the functions Δ and r to any vector $\boldsymbol{v} \in \mathbb{R}^m$ and any matrix $A \in \mathbb{R}^{m \times \lfloor H \rfloor}$, in a straightforward manner. Similarly, let us extend the functions θ and Θ to any matrix $A' \in \mathbb{R}^{m \times \lfloor H \rfloor}$.

5.2.1 Existence and Convexity of the Objective Function

Proposition 5 *The function* $\Delta(\boldsymbol{v}, A)$ *is convex.*

Proof. For any fixed $\boldsymbol{\phi}$, the function $\boldsymbol{q}^T [\, \boldsymbol{b} - A\boldsymbol{\phi} - \boldsymbol{v} \,]_+$ is clearly convex in $(\boldsymbol{v}, A)$. The expectation inherits convexity. □

Corollary 6 *The function* r *is convex.*

Proof. For arbitrary $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^m$, let $A_{\boldsymbol{v}}$ and $A_{\boldsymbol{w}}$ designate cutting rules that minimize $\Delta(\boldsymbol{v}, .)$ and $\Delta(\boldsymbol{w}, .)$, respectively. Let $0 \leq \lambda \leq 1$. From the definition and convexity of the inventory function, we have

$$\begin{aligned} \lambda r(\boldsymbol{v}) + (1-\lambda) r(\boldsymbol{w}) &= \lambda \Delta(\boldsymbol{v}, A_{\boldsymbol{v}}) + (1-\lambda)\Delta(\boldsymbol{w}, A_{\boldsymbol{w}}) \\ &\geq \Delta\Big(\lambda \boldsymbol{v} + (1-\lambda)\boldsymbol{w},\ \lambda A_{\boldsymbol{v}} + (1-\lambda) A_{\boldsymbol{w}} \Big) . \end{aligned}$$

Here $\lambda A_{\boldsymbol{v}} + (1-\lambda) A_{\boldsymbol{w}}$ is obviously a cutting rule. Hence the righ-hand-side expression is at least $r\Big(\lambda \boldsymbol{v} + (1-\lambda)\boldsymbol{w} \Big)$. □

Proposition 7 *The function r satisfies the following inequalities with any vectors $\boldsymbol{v}, \boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{x} \geq \boldsymbol{0}$:*

$$r(\,\boldsymbol{v}\,) + \boldsymbol{q}^T \boldsymbol{x} \;\geq\; r(\,\boldsymbol{v} - \boldsymbol{x}\,) \;\geq\; r(\,\boldsymbol{v}\,) \ .$$

Proof. Let $A_{\boldsymbol{v}}$ designate a cutting rule that minimizes $\Delta(\boldsymbol{v}, .)$. From the definition of the inventory function, we have

$$r(\,\boldsymbol{v}\,) \;=\; \mathrm{E}\Big(\boldsymbol{q}^T [\,\boldsymbol{b} - \boldsymbol{v} - A_{\boldsymbol{v}}\, \boldsymbol{\phi}\,]_+ \Big)$$

$$r(\,\boldsymbol{v} - \boldsymbol{x}\,) \;\leq\; \mathrm{E}\Big(\boldsymbol{q}^T [\,\boldsymbol{b} - \boldsymbol{v} + \boldsymbol{x} - A_{\boldsymbol{v}}\, \boldsymbol{\phi}\,]_+ \Big) \ .$$

Subtracting the equation from the inequality, we get

$$r(\,\boldsymbol{v} - \boldsymbol{x}\,) - r(\,\boldsymbol{v}\,) \;\leq\; \mathrm{E}\Big(\boldsymbol{q}^T [\,\boldsymbol{b} - \boldsymbol{v} + \boldsymbol{x} - A_{\boldsymbol{v}}\, \boldsymbol{\phi}\,]_+ \Big) - \mathrm{E}\Big(\boldsymbol{q}^T [\,\boldsymbol{b} - \boldsymbol{v} - A_{\boldsymbol{v}}\, \boldsymbol{\phi}\,]_+ \Big) \ .$$

For the first inequality of the proposition, it is enough to show that the right-hand side of the above inequality is less than or equal to $\boldsymbol{q}^T \boldsymbol{x}$. We have

$$[\,\boldsymbol{b} - \boldsymbol{v} + \boldsymbol{x} - A_{\boldsymbol{v}}\, \boldsymbol{\phi}\,]_+ - [\,\boldsymbol{b} - \boldsymbol{v} - A_{\boldsymbol{v}}\, \boldsymbol{\phi}\,]_+ \;\leq\; [\,\boldsymbol{x}\,]_+ \;=\; \boldsymbol{x} \ .$$

Form the scalar products of both sides with the non-negative vector $\boldsymbol{q}$, and take the expectations of both scalar products. Inequality is preserved by those operations.

The second inequality of the proposition (monotonicity of r) is obvious. □

Proposition 8 *For any fixed $\boldsymbol{\phi}'$, the function $\theta\,(\,A',\, \boldsymbol{\phi}'\,)$ is convex in A'.*

Actually, θ is a convex function of the vector $\boldsymbol{\eta} = A'\,\boldsymbol{\phi}' - \boldsymbol{b}'$.

Proof. Let $\lambda\,\boldsymbol{\eta} + (1-\lambda)\,\boldsymbol{\zeta}$ be a convex combination of the vectors $\boldsymbol{\eta}, \boldsymbol{\zeta} \in \mathbb{R}^m$. We show that the convexity inequality holds for those vectors. For the sake of simplicity, we consider only the case $\lambda = \frac{1}{2}$. Proof in the general case is similar.

The respective function values for the vectors $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ are the following

$$\begin{aligned} \boldsymbol{q}'^T [\,\boldsymbol{\eta}\,]_- \;+\; r(\,[\,\boldsymbol{\eta}\,]_+\,) \ , \\ \boldsymbol{q}'^T [\,\boldsymbol{\zeta}\,]_- \;+\; r(\,[\,\boldsymbol{\zeta}\,]_+\,) \ . \end{aligned} \tag{5.3}$$

Before going on with the formal proof, let us illustrate the idea. Imagine that the fiber producing factory consists of two identical production units. We compare two cases. In the first case, the units work independently, and each has to satisfy their respective demands separately. In the second case, an exchange of products is possible at the end of each production period.

The first-period demand is $\boldsymbol{b}'$ for each unit (the total demand being $2\boldsymbol{b}'$). For the first production unit, the vector $\boldsymbol{\eta}$ designates the numbers of surplus/lacking pieces in different products at the end of the first period. (A positive component represents a surplus, and a negative a shortage.) For the second unit, the vector $\boldsymbol{\zeta}$ designates the numbers of surplus/lacking products.

The numbers of the exchanged products are represented by the vector $\boldsymbol{x}$. Suppose that in a certain product length, there is a surplus in the first production unit, and a shortage in the second unit. Then some of the surplus products from the first unit are transferred to the second unit, until the shortage is made good. In the opposite case, transfer is made in the opposite direction. (Transfer in the opposite direction is expressed by a negative value.) Formally, for $k = 1, \ldots, m$, let

$$x_k = \begin{cases} & \min(\,\eta_k,\, -\zeta_k\,) & \text{if } \eta_k > 0 \text{ and } \zeta_k < 0 \\ - & \min(\,-\eta_k,\, \zeta_k\,) & \text{if } \eta_k < 0 \text{ and } \zeta_k > 0 \\ & 0 & \text{otherwise, i.e., if } \eta_k\zeta_k \geq 0 \ . \end{cases}$$

$x_k = 0$ means that $\eta_k\zeta_k \geq 0$ holds. This relation between η_k and ζ_k will be abbreviated as $\eta_k \sim \zeta_k$.

The vector $\boldsymbol{x}$ was defined in such a manner that $x_k \sim \eta_k - x_k$ holds for $k = 1, \ldots, m$. (Meaning that transfer increases neither the surplus, nor the shortage in the first unit.) This relation between vectors will be abbreviated as $\boldsymbol{x} \sim \boldsymbol{\eta} - \boldsymbol{x}$. From the definition of the vector $\boldsymbol{x}$, we also have $-\boldsymbol{x} \sim \boldsymbol{\zeta} + \boldsymbol{x}$ and $\boldsymbol{\eta} - \boldsymbol{x} \sim \boldsymbol{\zeta} + \boldsymbol{x}$.

From $\boldsymbol{x} \sim \boldsymbol{\eta} - \boldsymbol{x}$ and $-\boldsymbol{x} \sim \boldsymbol{\zeta} + \boldsymbol{x}$, we have respectively

$$[\,\boldsymbol{\eta}\,]_+ - [\,\boldsymbol{x}\,]_+ = [\,\boldsymbol{\eta} - \boldsymbol{x}\,]_+ \quad \text{and} \quad [\,\boldsymbol{\zeta}\,]_+ - [\,-\boldsymbol{x}\,]_+ = [\,\boldsymbol{\zeta} + \boldsymbol{x}\,]_+ \ .$$

Hence from Proposition 7, and the assumption $\boldsymbol{q}' \geq \boldsymbol{q}$, we have

$$\begin{aligned} r([\,\boldsymbol{\eta}\,]_+) \ + \ \boldsymbol{q}'^T [\,\boldsymbol{x}\,]_+ \ &\geq \ r([\,\boldsymbol{\eta} - \boldsymbol{x}\,]_+) \ , \\ r([\,\boldsymbol{\zeta}\,]_+) \ + \ \boldsymbol{q}'^T [\,\boldsymbol{x}\,]_- \ &\geq \ r([\,\boldsymbol{\zeta} + \boldsymbol{x}\,]_+) \ . \end{aligned} \tag{5.4}$$

Again, from $-\boldsymbol{x} \sim \boldsymbol{\zeta} + \boldsymbol{x}$ and $\boldsymbol{x} \sim \boldsymbol{\eta} - \boldsymbol{x}$, we have respectively

$$[\,-\boldsymbol{x}\,]_- = [\,\boldsymbol{\zeta}\,]_- - [\,\boldsymbol{\zeta} + \boldsymbol{x}\,]_- \quad \text{and} \quad [\,\boldsymbol{x}\,]_- = [\,\boldsymbol{\eta}\,]_- - [\,\boldsymbol{\eta} - \boldsymbol{x}\,]_-$$

Substituting these into the second terms of the respective left-hand sides of the inequalities (5.4), we get

$$\begin{aligned} \boldsymbol{q}'^T [\,\boldsymbol{\zeta}\,]_- \;+\; r([\,\boldsymbol{\eta}\,]_+) \;&\geq\; \boldsymbol{q}'^T [\,\boldsymbol{\zeta}+\boldsymbol{x}\,]_- \;+\; r([\,\boldsymbol{\eta}-\boldsymbol{x}\,]_+)\ , \\ \boldsymbol{q}'^T [\,\boldsymbol{\eta}\,]_- \;+\; r([\,\boldsymbol{\zeta}\,]_+) \;&\geq\; \boldsymbol{q}'^T [\,\boldsymbol{\eta}-\boldsymbol{x}\,]_- \;+\; r([\,\boldsymbol{\zeta}+\boldsymbol{x}\,]_+)\ . \end{aligned} \tag{5.5}$$

Add up the above two inequalities. From $\boldsymbol{\eta}-\boldsymbol{x} \sim \boldsymbol{\zeta}+\boldsymbol{x}$, it follows that

$$[\,\boldsymbol{\eta}-\boldsymbol{x}\,]_- + [\,\boldsymbol{\zeta}+\boldsymbol{x}\,]_- = [\,\boldsymbol{\eta}+\boldsymbol{\zeta}\,]_- \quad\text{and}\quad [\,\boldsymbol{\eta}-\boldsymbol{x}\,]_+ + [\,\boldsymbol{\zeta}+\boldsymbol{x}\,]_+ = [\,\boldsymbol{\eta}+\boldsymbol{\zeta}\,]_+\ .$$

Since the function r is convex, the sum of the right-hand sides of (5.5) is at least

$$2\,\boldsymbol{q}'^T \left[\,\frac{\boldsymbol{\eta}+\boldsymbol{\zeta}}{2}\,\right]_- \;+\; 2\,r\!\left(\left[\,\frac{\boldsymbol{\eta}+\boldsymbol{\zeta}}{2}\,\right]_+\right)\ . \tag{5.6}$$

From the pair of expressions (5.3), the pair of inequalities (5.5), and the expression (5.6), we get the convexity inequality. □

If the inventory $\boldsymbol{v}$ satisfies $v_k \geq b_k$ for the kth product ($1 \leq k \leq m$), then the demand for this product can be satisfied from the inventory. Such a product can simply be ignored when the second-period cutting rule is specified. This argument is formally stated in the following

Proposition 9 *For any vector* $\boldsymbol{v}$, *we have*

$$r(\,\boldsymbol{v}\,) \;=\; r\Big(\,\min(\boldsymbol{v},\boldsymbol{b})\,\Big)\ ,$$

where min *is meant componentwise.*

Proof. By definition,

$$r(\,\boldsymbol{v}\,) \;=\; \mathrm{E}\Big(\,\boldsymbol{q}^T\,[\,\boldsymbol{b}\,-\,\boldsymbol{v}\,-\,A_{\boldsymbol{v}}\,\boldsymbol{\phi}\,]_+\,\Big)\ ,$$

where $A_{\boldsymbol{v}}$ is a cutting rule which minimizes the expectation. Let us compute the expectation componentwise:

$$\sum_{k=1}^{m}\ \mathrm{E}\Big(\,q_k\,\Big[\,b_k\,-\,w_k\,-\,\boldsymbol{a}^k_{\boldsymbol{v}}\,\boldsymbol{\phi}\,\Big]_+\,\Big)\ ,$$

where $\boldsymbol{a}^k_{\boldsymbol{v}}$ is the kth row of the matrix $A_{\boldsymbol{v}}$.

Suppose $w_k \geq b_k$ holds for some $1 \leq k \leq m$. Then obviously there exists such optimal cutting rule $A_{\boldsymbol{v}}$, which satisfies $\boldsymbol{a}^k_{\boldsymbol{v}} = \boldsymbol{0}$. Hence the kth term of the above sum is zero. □

Theorem 10 *The expectation* $\Theta(A')$ *exists and is finite for any matrix* A'.

Proof. The function r is convex (Proposition 6), hence it is continuous. (About convexity implying continuity, see Theorem 10.1 in Rockafellar (1970).) It follows that the function θ is also continuous. Since it is also non-negative, only the finiteness of the expectation needs proof.

When computing $\theta(A', \boldsymbol{\phi}')$, the function value $r(\boldsymbol{v})$ must be evaluated only for vectors $\boldsymbol{v} \geq \mathbf{0}$. Proposition 9 asserts that the supremum of the function r over the positive orthant equals the supremum over the m-dimensional closed brick

$$B \;=\; [0, b_1] \times \cdots \times [0, b_m] \;\; .$$

The function r being continuous, and the brick compact, we have

$$r\left([A'\boldsymbol{\phi}' - \boldsymbol{b}']_+ \right) \;\leq\; R \quad \text{for some } R \in \mathbb{R} \;\; .$$

It follows that

$$\theta(A', \boldsymbol{\phi}') \;\leq\; \boldsymbol{q}'^T [A'\boldsymbol{\phi}' - \boldsymbol{b}']_- \;+\; R \;\; ,$$

and so

$$\Theta(A') \;\leq\; \mathrm{E}\Big(\boldsymbol{q}'^T [A'\boldsymbol{\phi}' - \boldsymbol{b}']_- \Big) \;+\; R \;\; . \tag{5.7}$$

(Expectation is considered with respect to the random vector $\boldsymbol{\phi}'$.)

For a fixed cutting rule $A' \in \mathbb{R}^{m \times \lfloor H \rfloor}$, the random vector $\boldsymbol{\eta} = A'\boldsymbol{\phi}' - \boldsymbol{b}'$ is normally distributed. The expectation on the right-hand side of inequality (5.7) must be split to the sum of 2^m terms, according to the orthants of the space $\mathbb{R}^m$. Each of the terms are clearly finite. □

Theorem 11 *The objective function Θ is convex.*

Proof. For any fixed $\boldsymbol{\phi}'$, the function $\theta(A', \boldsymbol{\phi}')$ is convex in A' (that is Proposition 8). Θ being the expectation of θ, it inherits convexity. □

5.2.2 Differentiability of the Inventory Function

In this section we show that subgradients of the inventory function can be computed by the solution of one-period problems.

Proposition 12 *For any subgradient $\boldsymbol{g}$ of the convex function r, we have*

$$\mathbf{0} \;\geq\; \boldsymbol{g} \;\geq\; -\boldsymbol{q} \;\; .$$

Proof. First we prove non-positivity of the subgradient. Let $\boldsymbol{g}$ be a subgradient at $\boldsymbol{v} - \boldsymbol{e}_k$. From the definition of the subgradient, we have

$$r(\boldsymbol{v}) \;\geq\; r(\boldsymbol{v} - \boldsymbol{e}_k) \;+\; \boldsymbol{g}^T \boldsymbol{e}_k$$

From Proposition 7, it follows that the left-hand side of this inequality is less than or equal to $r(\boldsymbol{v} - \boldsymbol{e}_k)$. Hence we have $0 \geq g_k$.

Consider now the second inequality of the proposition. Let $\boldsymbol{g}$ be a subgradient at $\boldsymbol{v}$. From the definition of the subgradient, we have

$$r(\boldsymbol{v} - \boldsymbol{e}_k) \;\geq\; r(\boldsymbol{v}) \;+\; \boldsymbol{g}^T(-\boldsymbol{e}_k) \;,$$

where $\boldsymbol{e}_k$ is the kth unit vector. From Proposition 7, it follows that the left-hand side of this inequality is less than or equal to

$$r(\boldsymbol{v}) \;+\; \boldsymbol{q}^T \boldsymbol{e}_k \;.$$

Hence we have $q_k \geq -g_k$. □

A direct consequence of Proposition 12 is that the function r satisfies the Lipschitz condition with the constant $\|\boldsymbol{q}\|$.

Now we examine differentiability. First let us fix a matrix A, and consider $\Delta(\boldsymbol{v}, A)$ as a function of $\boldsymbol{v}$. This function is convex (Proposition 5). We are going to compute its subgradients.

Consider a component $1 \leq k \leq m$, and denote the kth row of A by $\boldsymbol{a}^k$. For any fixed $\boldsymbol{a}^k$ and $\boldsymbol{\phi}$, the function $v_k \mapsto q_k \left[b_k - \boldsymbol{a}^k\boldsymbol{\phi} - v_k \right]_+$ is clearly convex and almost everywhere differentiable in v_k. For any v_k pick a subgradient of the above function, and denote it by $\gamma_k\left(v_k, \boldsymbol{a}^k, \boldsymbol{\phi}\right)$. That subgradient is also a derivative for almost every v_k and can be computed as

$$\gamma_k\left(v_k, \boldsymbol{a}^k, \boldsymbol{\phi}\right) \;=\; \begin{cases} 0 \;, & \text{if } b_k - \boldsymbol{a}^k\boldsymbol{\phi} \;<\; v_k \\ \\ -q_k \;, & \text{if } b_k - \boldsymbol{a}^k\boldsymbol{\phi} \;>\; v_k \;. \end{cases} \tag{5.8}$$

(The function is not differentiable in the case $b_k - \boldsymbol{a}^k\boldsymbol{\phi} = v_k$.) For fixed v_k and $\boldsymbol{a}^k$, the value $\gamma_k\left(v_k, \boldsymbol{a}^k, \boldsymbol{\phi}\right)$ is a random variable. The expectation of this variable clearly exists, and can be computed as

$$\Gamma_k\left(v_k, \boldsymbol{a}^k\right) \;=\; -q_k \,\mathrm{P}(\,\boldsymbol{a}^k\boldsymbol{\phi} \;<\; b_k - v_k\,) \;.$$

Proposition 13 *For any $m \times \lfloor H \rfloor$ matrix A, the m-vector*

$$\boldsymbol{\Gamma}(\,\boldsymbol{v}, A\,) \;=\; \left(\,\Gamma_1\left(v_1, \boldsymbol{a}^1\right), \;\ldots,\; \Gamma_m(\,v_m, \boldsymbol{a}^m\,)\,\right)$$

is a subgradient of the function $\boldsymbol{v} \mapsto \Delta(\boldsymbol{v}, A)$ at $\boldsymbol{v}$. Formally, the following inequality holds

$$\Delta(\boldsymbol{w}, A) \;\geq\; \Delta(\boldsymbol{v}, A) \;+\; \left[\,\boldsymbol{\Gamma}(\,\boldsymbol{v}, A\,)\,\right]^T(\,\boldsymbol{w} \;-\; \boldsymbol{v}\,) \tag{5.9}$$

for any vectors $\boldsymbol{w}$ and $\boldsymbol{v}$.

Proof. From the definition of $\gamma_k\left(v_k,\, \boldsymbol{a}^k,\, \boldsymbol{\phi}\right)$, we have

$$q_k\left[b_k - \boldsymbol{a}^k\boldsymbol{\phi} - w_k\right]_+ \geq q_k\left[b_k - \boldsymbol{a}^k\boldsymbol{\phi} - v_k\right]_+ + \left[\gamma_k\left(v_k,\, \boldsymbol{a}^k,\, \boldsymbol{\phi}\right)\right]^T (w_k - v_k)$$

for any realizations of the random vector $\boldsymbol{\phi}$, and for any component $k = 1, \ldots, m$. Taking expected values of both sides, and summing up according to k, we get (5.9). □

The subgradient $\Gamma_k\left(v_k,\, \boldsymbol{a}^k\right)$ is a continuous function of v_k, with the possible exception of $v_k = b_k$. (We may have $\boldsymbol{a}^k = \boldsymbol{0}$.) Hence $\Gamma_k\left(v_k,\, \boldsymbol{a}^k\right)$ is a unique subgradient of the function $v_k \mapsto \mathrm{E}\left(q_k\left[b_k - \boldsymbol{a}^k\boldsymbol{\phi} - v_k\right]_+\right)$ at v_k, provided $v_k \neq b_k$ holds.

It follows that for any fixed A, the function $\Delta(\boldsymbol{v},\, A)$ is differentiable in the first variable, provided $v_k \neq b_k$ $(k = 1, \ldots, m)$ holds. The gradient is $\boldsymbol{\Gamma}(\boldsymbol{v},\, A)$.

Now let us consider the function r. By definition, $r(\boldsymbol{v}) = \Delta(\boldsymbol{v},\, A_{\boldsymbol{v}})$, where $A_{\boldsymbol{v}} \in P$ is the cutting rule which minimizes $\Delta(\boldsymbol{v},\, .)$ over P. (That is, $\Delta(\boldsymbol{v},\, A_{\boldsymbol{v}}) \leq \Delta(\boldsymbol{v},\, A)$ holds for any $A \in P$.)

Proposition 14 *Let $\boldsymbol{v} \in \mathbb{R}^m$ be such an inventory vector that $v_k \neq b_k$ $(k = 1, \ldots, m)$ holds.*

Then the function r is differentiable at $\boldsymbol{v}$, and the gradient is $\boldsymbol{\Gamma}(\boldsymbol{v},\, A_{\boldsymbol{v}})$.

Proof. Let $\boldsymbol{g} \in \mathbb{R}^m$ be a subgradient of the function r at $\boldsymbol{v}$. We have

$$\Delta(\boldsymbol{w},\, A_{\boldsymbol{v}}) \geq r(\boldsymbol{w}) \geq r(\boldsymbol{v}) + \boldsymbol{g}^T(\boldsymbol{w} - \boldsymbol{v})$$

for any $\boldsymbol{w} \in \mathbb{R}^m$. (The first inequality follows from the definition of r, the second from the definition of $\boldsymbol{g}$.) Consider the expressions on the left- and right-hand sides of the above inequality as functions of $\boldsymbol{w}$. In the point $\boldsymbol{w} = \boldsymbol{v}$, the function values are equal. Moreover, both functions are differentiable at this point, and the respective gradients are $\boldsymbol{\Gamma}(\boldsymbol{v},\, A_{\boldsymbol{v}})$ and $\boldsymbol{g}$.

Hence $\boldsymbol{g} = \boldsymbol{\Gamma}(\boldsymbol{v},\, A_{\boldsymbol{v}})$ must hold. Uniqueness of the subgradient implies differentiability (see Theorem 25.1 in Rockafellar (1970)). □

If a convex function is differentiable on an open convex set, then it is actually smooth (continuously differentiable) on that set (see Corollary 25.5.1 in Rockafellar (1970)). Hence the inventory function is smooth on the set $\{\, \boldsymbol{v} \in \mathbb{R}^m \mid v_k \neq b_k \;\; (k = 1, \ldots, m) \,\}$. (Consider the above m-dimensional set as the union of 2^m open convex orthants.)

Remark. For a given non-negative vector $\boldsymbol{v} \in \mathbb{R}^m$, the optimal cutting rule $A_{\boldsymbol{v}}$ can be computed by the solution of a one-period problem. Given the matrix $A_{\boldsymbol{v}}$, the random vector $A_{\boldsymbol{v}}\boldsymbol{\phi}$ has a normal distribution of known parameters. Hence the gradient $\nabla r(\boldsymbol{v}) = \boldsymbol{\Gamma}(\boldsymbol{v},\, A_{\boldsymbol{v}})$ can be computed.

5.2.3 Evaluating the Inventory Function in the Positive Orthant

In order to compute the objective function, the inventory function needs to be evaluated only in the positive orthant $\mathbb{R}^m_+$. Moreover, we have $r(\boldsymbol{v}) = r(\min(\boldsymbol{v}, \boldsymbol{b}))$ for any vector $\boldsymbol{v}$ (Proposition 9). Hence the function r needs to be evaluated only in the brick

$$B = [0, b_1] \times \cdots \times [0, b_m] .$$

Evaluating the inventory function at a given point means the solution of a one-period problem. Some of the products may be ignored when solving the one-period problem: the products in which demand can be satisfied from stock. (The set of those products is $K = \{ k \mid k \in \{1, \ldots, m\},\ v_k \geq b_k \}$.)

The brick B is the disjoint union of the half-open facets

$$F_K = \prod_{k \in K} \{b_k\} \times \prod_{k \notin K} [0, b_k) , \qquad K \subset \{1, \ldots, m\} .$$

The evaluation of r in a point of the facet F_K requires the solution of a one-period problem of dimension $m - |K|$.

In the previous section we proved that the inventory function is smooth on the facet $F_\emptyset$. Consider restricting the inventory function to a lower-dimension facet. All the arguments presented in the previous section apply in lower dimensions as well.

Hence the restricted function is smooth on any facet F_K, and we can compute the gradient of the restricted function. From the gradient of the restricted function at $\boldsymbol{v} \in F_K$, we can construct a subgradient of the $\mathbb{R}^m \longrightarrow \mathbb{R}$ inventory function. (The restricted gradient vector must be lengthened by adding 0 components at positions $k \in K$.)

5.3 Lower Approximation of the Objective Function

We propose to estimate the objective function value by simulation (by repeated evaluation of the inventory function). In Section 5.2.3 we showed that the inventory function needs to be evaluated only in points of the brick B. The difficulty is that given a point $\boldsymbol{v} \in B$, the computation of the function value $r(\boldsymbol{v})$ requires the solution of a one-period problem. Hence the evaluation of the function for every random vector generated, would mean an excessive amount of work.

The aim of the present section is to overcome this difficulty. We approximate the inventory function with a quickly computable function.

Let us evaluate the inventory function at the points $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in B$. These points will be referred to as *basis points*. We can also compute a subgradient of the inventory function at each basis point. Let us denote the selected subgradients at the basis points by $\nabla r(\boldsymbol{v}_i),\ i = 1, \ldots, n$.

Remark 15 *For points in the boundary of the brick, let us construct the subgradient as described in Section 5.2.3. Hence for $\boldsymbol{v}_i \in F_K$, we have $[\nabla r(\boldsymbol{v}_i)]_k = 0 \quad (k \in K)$.*

Let us define the linear support functions $l_i : \mathbb{R}^m \longrightarrow \mathbb{R}$ $(i = 1, \ldots, n)$ as follows

$$l_i(\boldsymbol{v}) \;=\; r(\boldsymbol{v}_i) \;+\; \Big[\, \nabla r(\boldsymbol{v}_i) \,\Big]^T (\boldsymbol{v} - \boldsymbol{v}_i) \;\; .$$

From the convexity of r, we have $r \geq l_i$.
Moreover, let $l_0 : \mathbb{R}^m \longrightarrow \mathbb{R}$ be identically equal to zero. Obviously, we have $r \geq l_0$.

The approximating function $\underline{r}$ is defined as the upper cover of the linear functions:

$$\underline{r}(\boldsymbol{v}) \;\;=\;\; \max_{0 \leq i \leq n} \;\; l_i(\boldsymbol{v}) \;\; .$$

We have the relations

$$\underline{r}(\boldsymbol{v}_i) \;=\; r(\boldsymbol{v}_i) \quad (i = 1, \ldots, n) \;\; .$$

For any vector $\boldsymbol{v} \in \mathbb{R}^m$, we have

$$\underline{r}(\boldsymbol{v}) \;\leq\; r(\boldsymbol{v}) \; .$$

The function $\underline{r}$ is continuous, convex, and piecewise linear (polyhedral). This means that the space $\mathbb{R}^m$ can be subdivided into a finite number of convex polyhedra with disjoint interiors such that the function is linear over each polyhedron. In our case, the ith polyhedron is given by the relations $l_i \geq l_j$, $j \neq i$ and $\underline{r} = l_i$.

In the interior of the ith polyhedron $(1 \leq i \leq n)$, the gradient of $\underline{r}$ is $\nabla r(\boldsymbol{v}_i)$. In the interior of the 0th polyhedron, the gradient of $\underline{r}$ is $\mathbf{0}$.

The objective function is approximated as follows

$$\underline{\theta}(A', \boldsymbol{\phi}') \;\;=\;\; \boldsymbol{q}'^T [\, A' \boldsymbol{\phi}' \,-\, \boldsymbol{b}' \,]_- \;+\; \underline{r}\,(\,[\, A' \boldsymbol{\phi}' \,-\, \boldsymbol{b}' \,]_+\,) \;\; ,$$

$$\underline{\Theta}\,(A') \;\;=\;\; \mathrm{E}\Big(\, \underline{\theta}(A', \boldsymbol{\phi}') \,\Big) \;\; .$$

(Expectation is considered with respect to the random vector $\boldsymbol{\phi}'$.)

First we prove that the above expectation exists, and the function $\underline{\Theta}$ is convex. Then we show that the subgradients of the approximate objective function can be computed as expectations. Finally, we examine the accuracy of the lower approximation.

5.3.1 Existence and Convexity

Let us examine the functions $\underline{r}$ and $\underline{\theta}$.

Proposition 16 *For any subgradient g of the convex function $\underline{r}$, the following inequality holds*

$$\mathbf{0} \;\geq\; \boldsymbol{g} \;\geq\; -\boldsymbol{q} \;\; .$$

Proof. Any subgradient of the piecewise linear function $\underline{r}$ is the convex combination of the gradients of the linear functions l_i. □

A direct consequence of Proposition 16 is that the function $\underline{r}$ satisfies the Lipschitz condition with the constant $\|\boldsymbol{q}\|$.

Corollary 17 *The function $\underline{r}$ satisfies the following inequalities with any vectors $\boldsymbol{v}, \boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{x} \geq \boldsymbol{0}$*

$$\underline{r}(\boldsymbol{v}) + \boldsymbol{q}^T \boldsymbol{x} \geq \underline{r}(\boldsymbol{v} - \boldsymbol{x}) \geq \underline{r}(\boldsymbol{v}) .$$

Proof. Let $\boldsymbol{g}$ be a subgradient at $\boldsymbol{v} - \boldsymbol{x}$. From the definition of the subgradient, we have

$$\underline{r}(\boldsymbol{v}) \geq \underline{r}(\boldsymbol{v} - \boldsymbol{x}) + \boldsymbol{g}^T \boldsymbol{x} .$$

From Proposition 16, it follows that the right-hand side of the above inequality is greater than or equal to $\underline{r}(\boldsymbol{v} - \boldsymbol{x}) - \boldsymbol{q}^T \boldsymbol{x}$.

The other inequality can be proven similarly. □

Proposition 18 *For any fixed $\boldsymbol{\phi}'$, the function $\underline{\theta}(A', \boldsymbol{\phi}')$ is convex in A'.*

Actually, $\underline{\theta}$ is a convex function of the vector $A'\boldsymbol{\phi}' - \boldsymbol{b}'$.

Proof. The same as the proof of Proposition 8. (Use Corollary 17 instead of Proposition 7.) □

Theorem 19 *The expectation $\underline{\Theta}(A')$ exists and is finite for any matrix A'.*

Proof. The same as the proof of Theorem 10. (Use the inequality $r \geq \underline{r} \geq 0$ to find an upper bound on $\underline{r}$ over the m-dimensional closed brick B.) □

Theorem 20 *The approximate objective function $\underline{\Theta}$ is convex.*

Proof. Follows from Proposition 18. □

5.3.2 Subdifferentiability of the Approximate Objective Function

The approximate objective function was defined as the expected value of the function $\underline{\theta}$. The aim of this section is to show that the expectation vector of the subgradient of $\underline{\theta}$ is a subgradient of the approximate objective function.

Let $\boldsymbol{\phi}'$ be a fixed yield. The function

$$\underline{\theta}(\,.\,, \boldsymbol{\phi}') : \mathbb{R}^{m \times \lfloor H \rfloor} \longrightarrow \mathbb{R}$$

is the composition of the following functions

$$\begin{aligned} \boldsymbol{\mathcal{L}} \ : \ \mathbb{R}^{m\times\lfloor H\rfloor} &\longrightarrow \mathbb{R}^m \\ A' &\mapsto A'\,\boldsymbol{\phi}' \ - \ \boldsymbol{b}' \ , \end{aligned}$$

$$\begin{aligned} \text{and} \quad \mathcal{C} \ : \quad \mathbb{R}^m &\longrightarrow \mathbb{R} \\ \boldsymbol{\eta} &\mapsto \boldsymbol{q}'^T [\,\boldsymbol{\eta}\,]_- \ + \ \underline{r}(\,[\,\boldsymbol{\eta}\,]_+\,) \ . \end{aligned}$$

The function $\mathcal{C}$ is convex (Proposition 18). Let us pick a subgradient of $\mathcal{C}$ for each $\boldsymbol{\eta} \in \mathbb{R}^m$, and denote it by $\boldsymbol{\mathcal{G}}(\boldsymbol{\eta})$. We describe the construction of $\boldsymbol{\mathcal{G}}(\boldsymbol{\eta})$:

The space $\mathbb{R}^m$ can be covered with a finite number of convex polyhedra

$$R_1\,,\ \ldots\ R_n \ ,$$

in such a way that the function $\underline{r}$ is linear over each polyhedron. The gradient vectors of the linear functions are $\nabla r(\,\boldsymbol{v}_i\,)$ $(i = 1, \ldots, n)$.

Let us introduce the notation

$$\bar{R}_i \ = \ \left\{\, \boldsymbol{\eta} \in \mathbb{R}^m \;\middle|\; [\,\boldsymbol{\eta}\,]_+ \in R_i \,\right\} \qquad (i = 1, \ldots, n) \ .$$

Suppose $\boldsymbol{\eta}$ lies in the interior of $\bar{R}_i$. The kth component of the subgradient vector $\boldsymbol{\mathcal{G}}(\boldsymbol{\eta}) \in \mathbb{R}^m$ can be constructed as follows:

$$\mathcal{G}_k(\boldsymbol{\eta}) \ = \ \begin{cases} -q_k' & \text{if } \eta_k < 0 \\ \\ [\nabla r(\,\boldsymbol{v}_i\,)]_k & \text{if } \eta_k \geq 0 \end{cases} \qquad (k = 1, \ldots, m) \quad .$$

(On the common boundary of two or more polyhedra $\bar{R}_i$, the subgradient can be defined e.g. according to the polyhedron with the smallest subsrcipt i.)

Let us split the polyhedron $\bar{R}_i$ according to the orthants of the space $\mathbb{R}^m$. The above constructed subgradient is constant over each split polyhedron.

Now let us turn our attention to the linear function $\boldsymbol{\mathcal{L}}$. From the differentiability viewpoint, elements of $\text{Domain}(\boldsymbol{\mathcal{L}})$ will be considered vectors in $\mathbb{R}^{m\lfloor H\rfloor}$ rather than matrices in $\mathbb{R}^{m\times\lfloor H\rfloor}$. (Let us put the columns of the matrix one after the other.) The differential of $\boldsymbol{\mathcal{L}}$ will then be the $m \times m\lfloor H\rfloor$ matrix

$$\left(\, \phi_1{}'I,\ \ldots,\ \phi_{\lfloor H\rfloor}{}'I \,\right) \ ,$$

where I means the $m \times m$ identity matrix. It is easy to prove that the $m\lfloor H\rfloor$-vector

$$[\,\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{L}}(A')\,]^T \left(\, \phi_1{}'I,\ \ldots,\ \phi_{\lfloor H\rfloor}{}'I \,\right) \tag{5.10}$$

is a subgradient of the composite function $\mathcal{C}\boldsymbol{\mathcal{L}}$ at A'.

Let us consider the expectation of (5.10) with respect to the random vector $\boldsymbol{\phi}'$. Denote the expectation by $\underline{\boldsymbol{\Psi}}(\,A'\,)$

Theorem 21 *The expectation* $\underline{\boldsymbol{\Psi}}(A')$ *exists and is finite for any* A'.

Proof. Let us prove componentwise. The component $(h-1)m+k$ $(k=1,\dots,m;\ h=1,\dots,\lfloor H\rfloor)$ of the vector (5.10) is

$$\mathcal{G}_k(A'\boldsymbol{\phi}' - \boldsymbol{b}')\,\phi_h' \,. \tag{5.11}$$

From the construction of $\mathcal{G}_k$, we have $0 \geq \mathcal{G}_k(\boldsymbol{\eta}) \geq -q_k$. We prove that the expectation of the absolute value of (5.11) exists. We have

$$|\,\mathcal{G}_k(A'\boldsymbol{\phi}' - \boldsymbol{b}')\,\phi_h'\,| \;\leq\; q_k\,|\phi_h'|\,.$$

The expectation of the right-hand side is clearly finite. Moreover, it can be computed since ϕ_h' has normal distribution of known parameters. □

Theorem 22 *For any matrix* A', *the expectation* $\underline{\boldsymbol{\Psi}}(A')$ *is a subgradient of the function* $\underline{\Theta}$ *at* A'.

Proof. The vector (5.10) is a subgradient of the convex function $\mathcal{CL} = \underline{\theta}(\,.\,,\boldsymbol{\phi}')$ at A'. Consider the inequality that charcterizes the subgradient, and take the expectations (according to $\boldsymbol{\phi}'$) of both sides of the inequality. □

Remark. The subgradient $\underline{\boldsymbol{\Psi}}(A')$ is an $m\lfloor H\rfloor$-vector. The absolute value of the component $(h-1)m+k$ is not greater then $q_k\,\mathrm{E}(|\phi_h'|)$.

5.3.3 Evaluating the Approximate Inventory Function in the Positive Orthant

This section is the analog of Section 5.2.3: in order to make the approximation $\underline{r}$ accurate in the positive orthant, it is enough to ensure that it is accurate in the facets $K \subset \{1,\dots,m\}$ of the brick B.

Given $K \subset \{1,\dots,m\}$, let $\underline{r}_K$ designate the upper cover of some of the linear support functions:

$$\underline{r}_K : \mathbb{R}^m \longrightarrow \mathbb{R}\,, \qquad \underline{r}_K \;=\; \max\{\, l_i \;\mid\; 1 \leq i \leq n,\;\; \boldsymbol{v}_i \in \mathrm{cl}(F_K)\,\}\,,$$

where $\mathrm{cl}(F_K)$ is the closure of the half-open facet F_K.

$\underline{r}_K$ is convex and piecewise linear. Moreover, for any subgradient $\boldsymbol{g}$, we have $\mathbf{0} \geq \boldsymbol{g} \geq -\boldsymbol{q}$. It follows that $\underline{r}_K$ satisfies the Lipschitz condition with the constant $\|\boldsymbol{q}\|$.

Proposition 23 *Given* $\boldsymbol{v} \in \mathbb{R}^m_+$, *let* $\boldsymbol{w} = \min(\boldsymbol{v},\boldsymbol{b})$ *and* $K = \{\,k \,|\, k \in \{1,\dots,m\},\, v_k \geq b_k\,\}$. *Then we have* $\underline{r}_K(\boldsymbol{w}) \leq \underline{r}(\boldsymbol{v})$.

Proof. $\boldsymbol{w}$ falls into F_K, and the subgradients $\nabla r(\boldsymbol{v}_i)$ $(\boldsymbol{v}_i \in \mathrm{cl}(F_K))$ were constructed in such a way that $[\nabla r(\boldsymbol{v}_i)]_k = 0$ $(k \in K)$ holds. (According to the Remark 15 at the beginning of Section 5.3). Hence we have $\underline{r}_K(\boldsymbol{w}) = \underline{r}_K(\boldsymbol{v})$.

The proposition follows from the obvious fact $\underline{r}_K \leq \underline{r}$. □

5.4 Computing Upper Bounds for the Objective Value

Upper bounds will be used to measure accuracy of the lower approximation presented in Section 5.3. Given a cutting rule A', we propose to compute an upper bound for the objective value $\Theta(A')$ by defining an upper approximation for the inventory function.

For a vector $\boldsymbol{v} \in \mathbb{R}^m$, let us define the upper approximation $\bar{r}(\boldsymbol{v})$ as the optimal objective value of the following pair of linear programming problems

$$\begin{array}{lll}
\mathrm{LP}(\boldsymbol{v}): & \text{find} & \boldsymbol{g} \in \mathbb{R}^m,\ h \in \mathbb{R} \\
& & \mathbf{0} \geq \boldsymbol{g} \geq -\boldsymbol{q} \\
& & \boldsymbol{v}_i{}^T \boldsymbol{g} + h \leq r(\boldsymbol{v}_i) \qquad (i = 1, \ldots, n) \\
& \max & \boldsymbol{v}^T \boldsymbol{g} + h \ . \\
\mathrm{DUAL}(\boldsymbol{v}): & \text{find} & \boldsymbol{\lambda} \in \mathbb{R}^n,\ \boldsymbol{\nu} \in \mathbb{R}^m \\
& & \boldsymbol{\lambda} \geq \mathbf{0},\ \boldsymbol{\nu} \geq \mathbf{0} \\
& & \sum_{i=1}^{n} \lambda_i = 1 \\
& & \sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i - \boldsymbol{\nu} \leq \boldsymbol{v} \\
& \min & \sum_{i=1}^{n} \lambda_i r(\boldsymbol{v}_i) + \boldsymbol{\nu}^T \boldsymbol{q} \ .
\end{array}$$

Note that $\mathrm{DUAL}(\boldsymbol{v})$ is the dual of $\mathrm{LP}(\boldsymbol{v})$. It is easy to check that both problems have a feasible solution for any $\boldsymbol{v}$. (Choose h very small and $\boldsymbol{\nu}$ very large.) From the duality theorem follows that both problems have optimal solutions, and their respective optimal objective values are equal.

Proposition 24 *We have the relations*

$$\bar{r}(\boldsymbol{v}_i) = r(\boldsymbol{v}_i) \quad (i = 1, \ldots, n) \ .$$

For any vector $\boldsymbol{v} \in \mathbb{R}^m$, *we have*

$$\bar{r}(\boldsymbol{v}) \geq r(\boldsymbol{v}) \ .$$

Proof. Let us consider the problem $\mathrm{LP}(\boldsymbol{v}_i)$, and choose the variables as follows. Let $\boldsymbol{g}$ be a subgradient of the convex function r at $\boldsymbol{v}_i$. Let $h = r(\boldsymbol{v}_i) - \boldsymbol{g}^T \boldsymbol{v}_i$. Proposition 12 asserts that $\boldsymbol{g}$ is between the lower and upper bounds prescribed $\mathrm{LP}(\boldsymbol{v}_i)$. From the definition of the subgradient, we have

$$r(\boldsymbol{v}_j) \geq r(\boldsymbol{v}_i) + \boldsymbol{g}^T(\boldsymbol{v}_j - \boldsymbol{v}_i) = h + \boldsymbol{g}^T \boldsymbol{v}_j \qquad (j = 1, \ldots, n) \ .$$

Hence the above solution $(\boldsymbol{g}, h)$ is feasible. The objective value is $r(\boldsymbol{v}_i)$.

That objective value can not be increased, since we have the constraint

$$\boldsymbol{v}_i{}^T\boldsymbol{g} + h \leq r(\boldsymbol{v}_i) .$$

This proves the first part of the proposition.

Now we prove the second part. Let $(\lambda, \boldsymbol{\mu})$ be a feasible solution of DUAL$(\boldsymbol{v})$, and consider the objective value. From the convexity of r, we have

$$\sum_{i=1}^{n} \lambda_i\, r(\boldsymbol{v}_i) + \boldsymbol{\nu}^T\boldsymbol{q} \geq r\left(\sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i\right) + \boldsymbol{\nu}^T\boldsymbol{q} .$$

Since $\boldsymbol{\nu} \geq \mathbf{0}$ holds, it follows from the first part of Proposition 7 that

$$r\left(\sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i\right) + \boldsymbol{q}^T\boldsymbol{\nu} \geq r\left(\sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i - \boldsymbol{\nu}\right) .$$

Since $\sum_{i=1}^{n} \lambda_i\, \boldsymbol{v}_i - \boldsymbol{\nu} \leq \boldsymbol{v}$ holds, it follows from the monotonicity of r (second part of Proposition 7) that

$$r\left(\sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i - \boldsymbol{\nu}\right) \geq r(\boldsymbol{v}) .$$

Putting together the above three inequalities, we get the second part of the proposition. □

Proposition 25 *The upper approximating function $\bar{r}$ is continuous, convex, and piecewise linear. For any subgradient g of $\bar{r}$, the following inequality holds*

$$\mathbf{0} \geq \boldsymbol{g} \geq -\boldsymbol{q} .$$

Proof. As $\boldsymbol{v}$ varies, the problem LP$(\boldsymbol{v})$ can be considered a parametric programming problem. Those objective vectors for which a certain basis is optimal form a convex polyhedron. There is such a polyhedron for each feasible basis. While the objective vector remains in one of those polyhedra, the optimal solution $(\boldsymbol{g}, h)$ remains unchanged. Hence the optimal objective value $\boldsymbol{v}^T\boldsymbol{g} + h$ is a linear function of $\boldsymbol{v}$. □

A direct consequence of the above proposition is that the function $\bar{r}$ satisfies the Lipschitz condition with the constant $\|\boldsymbol{q}\|$.

Proposition 26 *For any vector $\boldsymbol{v}$, we have*

$$\bar{r}(\boldsymbol{v}) = \bar{r}\Big(\min(\boldsymbol{v}, \boldsymbol{b})\Big) .$$

Proof. Suppose the kth component of $\boldsymbol{v}$ is greater then b_k for some $k \in 1, \ldots, m$. Since the basis points are selected from the brick B, the kth component of each basis point at most b_k. Hence the kth component of the convex combination vector $\sum_{i=1}^{n} \lambda_i\, \boldsymbol{v}_i$ is less then the kth component of $\boldsymbol{v}$.

It follows that for the optimal solution $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ of DUAL$(\boldsymbol{v})$, we have $\nu_k = 0$. Hence decreasing the kth component of $\boldsymbol{v}$ will not affect the optimal objective value. □

Chapter 6

Estimation of the Objective Function Value

To implement a solution method for the dinamic problem, we must be able to estimate the objective value at any given point $A' \in P$.

Given a tolerance $\delta > 0$ and a probability $1 > p > 0$, we need random variables $\bar{\zeta}$ and $\underline{\zeta}$ such that

$$\mathrm{P}\left(\bar{\zeta} \geq \Theta(A') \text{ and } \Theta(A') \geq \underline{\zeta} \right) \geq p, \tag{6.1}$$

and $\bar{\zeta} - \underline{\zeta} \leq \delta$.

Moreover, realizations of the random variables $\bar{\zeta}$ and $\underline{\zeta}$ must be easy to generate.

The objective value $\Theta(A')$, is the expectation of the expression

$$\boldsymbol{q}'^T [\, A' \boldsymbol{\phi}' - \boldsymbol{b}' \,]_- + r\,([\, A' \boldsymbol{\phi}' - \boldsymbol{b}' \,]_+) .$$

By definition, expectation is considered with respect to the random vector $\boldsymbol{\phi}'$ which has a non-degenerate $\lfloor H \rfloor$-variate normal distribution with expectation vector $\boldsymbol{\mu}$ and covariance matrix C.

Instead of computing the expectation with respect to $\boldsymbol{\phi}'$, it is sufficient to compute it with respect to the significantly shorter random vector $\boldsymbol{\eta} = A'\boldsymbol{\phi}' - \boldsymbol{b}'$. For a fixed cutting rule A', the random vector $\boldsymbol{\eta}$ will also be normally distributed with expectation vector $A'\boldsymbol{\mu} - \boldsymbol{b}'$ and covariance matrix $A'CA'^T$.

The expectation of the first term of the above expression, i.e., of $\boldsymbol{q}'^T [\, \boldsymbol{\eta} \,]_-$, can be computed by the integration of the univariate normal probability density function. (The computation is described in Section 3.3.)

For the estimation of the expectation of the second term, the following statistical function will be used:

$$\frac{1}{M} \sum_{j=1}^{M} r([\boldsymbol{\eta}^{(j)}]_+) ,$$

where $\boldsymbol{\eta}^{(j)}$ $(j = 1, \ldots, M)$ designate mutually independent random vectors that have the same distribution as $\boldsymbol{\eta}$.

Let ϵ be a positive tolerance which is significantly smaller than the prescribed δ. In Section 6.1 we compute the sample size M which ensures that

$$\left| \frac{1}{M} \sum_{j=1}^{M} r([\boldsymbol{\eta}^{(j)}]_+) - \mathrm{E}\left\{ r([\boldsymbol{\eta}]_+) \right\} \right| < \epsilon \tag{6.2}$$

hold with probability p. This means that the random variables

$$\bar{\zeta} = \mathrm{E}(\boldsymbol{q}'^T [\boldsymbol{\eta}]_-) + \frac{1}{M} \sum_{j=1}^{M} \bar{r}([\boldsymbol{\eta}^{(j)}]_+) + \epsilon \quad \text{and} \quad \underline{\zeta} = \mathrm{E}(\boldsymbol{q}'^T [\boldsymbol{\eta}]_-) + \frac{1}{M} \sum_{j=1}^{M} \underline{r}([\boldsymbol{\eta}^{(j)}]_+) - \epsilon$$

satisfy

$$\mathrm{P}\left(\bar{\zeta} \geq \Theta(A') \text{ and } \Theta(A') \geq \underline{\zeta} \right) \geq p .$$

In order to generate realizations of the random variables $\bar{\zeta}$ and $\underline{\zeta}$ we need to generate realizations $\boldsymbol{y}^{(j)}$ $(j = 1, \ldots, M)$ of the random vector $\boldsymbol{\eta}$. A variety of methods for the generation of normally distributed random vectors can be found in Deák (1990).

The definition of the random variables $\bar{\zeta}$ and $\underline{\zeta}$ is not yet complete: they depend on the upper and lower approximations, respectively, of the inventory function. I.e., $\bar{\zeta}$ and $\underline{\zeta}$ depend on the set of the basis points $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. However, the probability requirement (6.1) is satisfied for any set of basis points. (The sample size M determined in Section 6.1 does not depend on the set of the basis points.)

In Section 6.2, we describe a procedure for the selection of the basis points $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. The selection depends on the realizations $\boldsymbol{y}^{(j)}$ $(j = 1, \ldots, M)$ of the random vector $\boldsymbol{\eta}$. Given the realizations $\boldsymbol{y}^{(j)}$, the basis points are selected in such a way that

$$\frac{1}{M} \sum_{j=1}^{M} \left\{ \bar{r}([\boldsymbol{y}^{(j)}]_+) - \underline{r}([\boldsymbol{y}^{(j)}]_+) \right\} \leq \delta - 2\epsilon . \tag{6.3}$$

will hold. (It could be done by including all the points $[\boldsymbol{y}^{(j)}]_+$ $(j = 1, \ldots, M)$ into the set of basis points. But the aim is to keep the number of the basis points small.)

Consider the procedure for the selection of the basis points to be part of the definition of the random variables $\bar{\zeta}$ and $\underline{\zeta}$. I.e., $\bar{\zeta}$ and $\underline{\zeta}$ are defined by means of a procedure that presents their respective realizations as functions of the realizations $\boldsymbol{y}^{(j)}$ $(j = 1, \ldots, M)$.

As a consequence of (6.3), we have $\bar{\zeta} - \underline{\zeta} \leq \delta$ as required.

6.1 Computation of the Sample Size

The size M of the sample required to achieve the prescribed accuracy and reliability (6.2) can be computed by using the dimensional central limit theorem. Let μ and σ^2 designate the expectation and variance, respectively, of the random variable $r([\boldsymbol{\eta}]_+)$.

We know that the expectation exists and is finite. First we prove that the variance is finite.

Proposition 27 *We have*

$$\sigma^2 \;\leq\; \|\boldsymbol{q}\|^2 \sum_{k=1}^{m} d_{kk} \; .$$

where d_{kk}, $k = 1, \ldots, m$ *are the diagonal elements of the covariance matrix* $A'CA'^T$ *of the random vector* $\boldsymbol{\eta}$.

Proof. For any random variable ξ, the expression $\mathrm{E}\,[(\xi - x)^2]$ is minimized by $x = E(\xi)$. Hence we have

$$\sigma^2 \;=\; \mathrm{E}\left[\left\{ r([\boldsymbol{\eta}]_+) \;-\; \mu \right\}^2 \right] \;\leq\; \mathrm{E}\left[\left\{ r([\boldsymbol{\eta}]_+) \;-\; r([\boldsymbol{\mu}]_+) \right\}^2 \right] \; . \tag{6.4}$$

Concerning the expression in the $\{\}$ brackets on the right-hand side of (6.4), we have

$$\left| r([\boldsymbol{\eta}]_+) \;-\; r([\boldsymbol{\mu}]_+) \right| \;\leq\; \|\boldsymbol{q}\| \left\| [\boldsymbol{\eta}]_+ \;-\; [\boldsymbol{\mu}]_+ \right\| \;\leq\; \|\boldsymbol{q}\| \, \| \boldsymbol{\eta} \;-\; \boldsymbol{\mu} \| \; ,$$

because the function r satisfies the Lipschitz condition with the constant $\|\boldsymbol{q}\|$. By substituting the above inequality into (6.4), we get

$$\sigma^2 \;\leq\; \mathrm{E}\left[\left\{ r([\boldsymbol{\eta}]_+) \;-\; r([\boldsymbol{\mu}]_+) \right\}^2 \right] \;\leq\; \|\boldsymbol{q}\|^2 \, \mathrm{E}\left(\| \boldsymbol{\eta} \;-\; \boldsymbol{\mu} \|^2 \right) \; .$$

Concerning the expectation on the right-hand side, we have

$$\mathrm{E}\left(\| \boldsymbol{\eta} \;-\; \boldsymbol{\mu} \|^2 \right) \;=\; \mathrm{E}\left(\sum_{k=1}^{m} (\eta_k - \mu_k)^2 \right) \;=\; \sum_{k=1}^{m} \mathrm{E}\left((\eta_k - \mu_k)^2 \right) \;=\; \sum_{k=1}^{m} d_{kk} \; .$$

□

The central limit theorem is applicable. It states that the distribution of the random variable

$$\frac{1}{M} \sum_{j=1}^{M} r(\, [\boldsymbol{\eta}^{(j)}]_+ \,)$$

can be approximated by a normal distribution, assumed that M is large enough. This normal distribution has the expectation μ and variance $\frac{1}{M}\sigma^2$.

Hence we have the following lower bound for the probability that the requirement (6.2) is satisfied:

$$\int_{-\epsilon}^{\epsilon} \frac{\sqrt{M}}{\sqrt{2\pi}\;\sigma} \; e^{-\frac{1}{2}\,Mx^2/\sigma^2} \, dx \;\;=\;\; \int_{-\epsilon\sqrt{M}/\sigma}^{\epsilon\sqrt{M}/\sigma} \frac{1}{\sqrt{2\pi}} \; e^{-\frac{1}{2}\,y^2} \, dy \;\;=\;\; 2\,\Phi\left(\frac{\epsilon\sqrt{M}}{\sigma} \right) - 1 \; ,$$

where Φ is the distribution function of the $N(0,1)$ -distribution. Let us introduce the notation $s = \|\boldsymbol{q}\| \sqrt{\sum_{k=1}^{m} d_{kk}}$. From Proposition 27, and the monotonicity of Φ, it follows that

$$2\,\Phi\left(\frac{\epsilon\sqrt{M}}{s} \right) - 1$$

is a lower bound for the probability that the requirement (6.2) is satisfied. Let us select the sample size M such that the above expression be at least p.

6.2 A Practical Approximation Scheme

In this section a heuristic method is described for the selection of the basis points. Given are the realizations $\boldsymbol{y}^{(j)}$ $(j = 1, \ldots, M)$ of the random vector $\boldsymbol{\eta}$. The basis points must be selected in such a way that (6.3) will hold. This will be achieved by a step-by-step extension of the set of the basis points. As a starting set, let us select the vertices of the brick B. New basis points will be added until (6.3) is satisfied. The aim is to keep the number of the basis points small.

Let us introduce the notation

$$\boldsymbol{z}^{(j)} \;=\; \min\Big([\boldsymbol{y}^{(j)}]_+ \,,\; \boldsymbol{b} \Big) \qquad (j = 1, \ldots, M)\,.$$

These vectors fall into the brick B. In Section 5.2.3, the brick was considered as the disjoint union of the facets F_K $(K \subset \{1, \ldots, m\})$. Hence given j $(1 \le j \le M)$, there exists a unique subset $K^{(j)} \subset \{1, \ldots, m\}$ for which $\boldsymbol{z}^{(j)} \in F_{K^{(j)}}$ holds. In Section 5.3.3, we defined the function $\underline{r}_{K^{(j)}}$ that is a lower approximation of the inventory function over the facet $F_{K^{(j)}}$.

From Propositions 26 and 23, respectively, we have

$$\bar{r}(\,[\boldsymbol{y}^{(j)}]_+\,) \;=\; \bar{r}(\,\boldsymbol{z}^{(j)}\,) \qquad \text{and} \qquad \underline{r}(\,[\boldsymbol{y}^{(j)}]_+\,) \;\ge\; \underline{r}_{K^{(j)}}(\,\boldsymbol{z}^{(j)}\,)\,.$$

It follows that

$$\bar{r}(\,[\boldsymbol{y}^{(j)}]_+\,) \;-\; \underline{r}(\,[\boldsymbol{y}^{(j)}]_+\,) \;\le\; \bar{r}(\,\boldsymbol{z}^{(j)}\,) \;-\; \underline{r}_{K^{(j)}}(\,\boldsymbol{z}^{(j)}\,)\,.$$

Let us introduce the notation

$$D^{(j)} \;=\; \bar{r}(\,\boldsymbol{z}^{(j)}\,) \;-\; \underline{r}_{K^{(j)}}(\,\boldsymbol{z}^{(j)}\,) \qquad (\,j = 1, \ldots, M\,)\,,$$

and consider the average

$$d \;=\; \frac{1}{M} \sum_{j=1}^{M} D^{(j)}\,.$$

We can ensure (6.3) by enforcing $d \;\le\; \delta - 2\epsilon$.

Sketch of the Procedure.

Initialization

At the start we have $n = 2^m$ basis points, the vertices of the brick B. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ designate these vertices.

The brick B is partitioned into 2^m disjoint rectangles (the facets F_K). Let $\mathcal{R}$ designate the set of these rectangles.

Estimation of Accuracy

The basis points determine the functions $\bar{r}$ and $\underline{r}_K$ $(K \subset \{1, \ldots, m\})$. Let us compute the differences $D^{(j)}$ $(j = 1, \ldots, M)$, and the average d.

If we have $d \leq \delta - 2\epsilon$ then (6.3) is satisfied, stop.

Extending the Set of the Basis Points

Let M_R $(R \in \mathcal{R})$ designate the number of those points $\boldsymbol{z}^{(j)}$ which fall into the rectangle R. (We have $\sum M_R = M$.) Let us introduce the notations

$$d_R = \frac{1}{M_R} \sum_{\boldsymbol{z}^{(j)} \in R} D^{(j)} \qquad (R \in \mathcal{R}) .$$

(If we have $M_R = 0$ for some R, then let $d_R = 0$.)

Let us find the rectangle $R^\star \in \mathcal{R}$ for which $\big(d_R - (\delta - 2\epsilon)\big) M_R$ is largest. Let $\boldsymbol{v}^\star$ denote the central point of $R^\star$. Let us subdivide the rectangle $R^\star$ into the disjoint union of sub-rectangles in such a way, that all the sub-rectangles have $\boldsymbol{v}^\star$ as a vertex. (If $R^\star$ is k-dimensional, then the number of sub-rectangles is 2^k. Details of the subdivision are illustrated by an example later in this section.)

Let us add the point $\boldsymbol{v}^\star$ to the set of basis points. (I.e., increment n, and let $\boldsymbol{v}_n = \boldsymbol{v}^\star$.) Let us delete $R^\star$ from the set $\mathcal{R}$, and include the sub-rectangles into which $R^\star$ was subdivided.

Go to step *Estimation of Accuracy*.

Example for the Subdivision of a Rectangle. For the sake of simplicity, assume that $R^\star$ is one of the facets of B, namely $R^\star = F_\emptyset$. The central point $\boldsymbol{v}^\star$ is then $\frac{1}{2}\boldsymbol{b}$.

$R^\star$ is an m-dimensional rectangle. It will be subdivided into the disjoint union of 2^m sub-rectangles. The construction of the sub-rectangles is the following: for any $L \subset \{1, \ldots, m\}$, let

$$\prod_{\ell \in L} [\, 0,\, \frac{1}{2} b_\ell \,) \;\times\; \prod_{\ell \notin L} [\, \frac{1}{2} b_\ell,\, b_\ell \,) \,.$$

Technical Remarks. Evaluation of the function $\underline{r}_K$ at a certain point requires the computation of n linear expressions, and selecting the largest of them.

Evaluation of the function $\bar{r}$, on the other hand, requires the solution of a linear programming problem, the matrix of which is of size $m \times n$. In our case m is small, hence the computation is possible even for a reasonably large n.

When recomputing the differences $D^{(j)}$ after $\bar{r}$ and $\underline{r}_K$ was redefined, results of former computations can be utilized.

Finiteness of the Procedure. Assume that a rectangle $R \in \mathcal{R}$ has a small diameter, namely,

$$\mathrm{Diam}(R) \;\leq\; \frac{\delta - 2\epsilon}{2\|\boldsymbol{q}\|}\,.$$

We prove that this rectangle will never be subdivided.

At least one vertex of R is a basis point. Let $\boldsymbol{v}$ designate such a vertex. Moreover, R was obtained by the (repeated) subdivision of one of the facets of the brick B. Let F_K designate this facet.

We have $\bar{r}(\boldsymbol{v}) = \underline{r}_K(\boldsymbol{v})$. Moreover, both $\bar{r}$ and $\underline{r}_K$ satisfy the Lipschitz condition with the constant $\|\boldsymbol{q}\|$. It follows that $D^{(j)} \leq 2\|\boldsymbol{q}\|\,\mathrm{Diam}(R) \leq \delta - 2\epsilon$ for any $\boldsymbol{z}^{(j)} \in R$. Hence we have $d_R \leq \delta - 2\epsilon$, and so $\Big(d_R - (\delta - 2\epsilon)\Big) M_R = 0$. This means that R will never be selected for subdivision. (If we have no better choice for $R^\star$, then the required precision has alreday been achieved.)

In each step of the procedure, a rectangle $R^\star$ is deleted from $\mathcal{R}$, and sub-rectangles are added. Each sub-rectangle has the diameter $\frac{1}{2}\mathrm{Diam}(R^\star)$. The number of the sub-rectangles is 2^k, where k is the dimension of $R^\star$. The procedure is bound to stop when every rectangle in $\mathcal{R}$ is small enough. It is easy to find a theoretical upper bound on the possible number of iterations.

Consider now the worst case, when this theoretical upper bound is reached. It means that for every $R \in \mathcal{R}$, we have $\mathrm{Diam}(R) \leq \frac{\delta - 2\epsilon}{2\|\boldsymbol{q}\|}$. The worst-case set $\mathcal{W}$ of the basis points is the widest possible. I.e., we have $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subset \mathcal{W}$ for any terminal set of basis points. Hence we have finitely many choices for the terminal set of basis points.

Chapter 7

Construction of Supporting Linear Functions

To implement a cutting planes method for the dinamic problem, we should be able to generate a supporting hyperplane for the objective function at any given point A'. We can not do that, but we have a good substitute: the linear support function of $\underline{\Theta}$ at A' clearly underestimates the objective function. The task is two-fold:

- The basis points $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ of the approximation must be selected in a way that ensures some prescribed accuracy. (I.e., at A', the difference between the objective value and the linear function value must not be greater than some prescribed tolerance. An upper estimator for $\Theta(A')$ can be derived from $\bar{r}$.)

- Since we can not compute exactly the value or gradient of $\underline{\Theta}$, we must estimate them by simulation.

The underestimating linear function is formally described as a *random linear function.* By this we mean a linear function, the coefficients of which are components of a random vector; and the constant of which is a random variable. The task is then:

Given a tolerance $\delta > 0$, and a probability $1 > p > 0$, we need a random linear function Λ such that

$$\Theta(A) \;\geq\; \Lambda(A) \quad (A \in P) \qquad \textit{and} \qquad \Theta(A') - \Lambda(A') \;\leq\; \delta \tag{7.1}$$

holds simultaneously with probability p. Moreover, relizations of the random linear function must be easy to generate.

Let $\delta_1, \delta_2 > 0$ be tolerances such that $\delta_1 + \delta_2 = \delta$. Moreover, let $1 > p_1, p_2 > 0$ be probabilities such that $p_1 p_2 = p$. Let us introduce the notations

$$\tilde{\delta_2} = \frac{\delta_2}{\sqrt{m\lfloor H \rfloor}\,\mathrm{Diam}(P)}\,, \qquad \tilde{p_2} = p_2\,,$$

where $\mathrm{Diam}(P)$ is the diameter of the feasible region. The task is accomplished in two steps.

Step 1 : Construct the random variables $\bar{\zeta}$ and $\underline{\zeta}$ as described in Section 7.1. The construction is such that

$$\mathrm{P}\left(\bar{\zeta} \geq \Theta(A') \text{ and } \underline{\Theta}(A') \geq \underline{\zeta} \right) \geq p_1 , \qquad \bar{\zeta} - \underline{\zeta} \leq \delta_1 .$$

The selection of appropriate basis points $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ for the approximation $\underline{\Theta}$ is part of the construction.

Step 2 : Construct the random $m\lfloor H \rfloor$-vector $\boldsymbol{\pi}$ as described in Section 7.2. Given a set of basis points, the construction is such that

$$\mathrm{P}\left(\| \boldsymbol{\pi} - \underline{\boldsymbol{\Psi}}(A') \|_{\max} \leq \tilde{\delta_2} \right) \geq \tilde{p_2} ,$$

where the approximation $\underline{\Theta}$ is based on the given set of points, and the $m\lfloor H \rfloor$-vector $\underline{\boldsymbol{\Psi}}(A')$ is the subgradient of $\underline{\Theta}$ at A'.

As input for the second step, we use the set of basis points $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ determined in the first step. Hence the following requirements are simultaneously satisfied with probability $p_1 \tilde{p_2} = p$:

$$\bar{\zeta} \geq \Theta(A') , \quad \underline{\Theta}(A') \geq \underline{\zeta} , \quad \bar{\zeta} - \underline{\zeta} \leq \delta_1 , \tag{7.2}$$

$$\| \boldsymbol{\pi} - \underline{\boldsymbol{\Psi}}(A') \|_{\max} \leq \tilde{\delta_2} . \tag{7.3}$$

The construction of the random linear function is as follows. In this respect cutting rules will be considered vectors in $\mathbb{R}^{m\lfloor H \rfloor}$ rather than matrices in $\mathbb{R}^{m \times \lfloor H \rfloor}$: Let us put the columns of the matrix one after the other. The initially given cutting rule will be denoted by $\boldsymbol{a}'$ rather than by A'. Let

$$\Lambda(\boldsymbol{a}) = (\underline{\zeta} - \delta_2) + \boldsymbol{\pi}^T (\boldsymbol{a} - \boldsymbol{a}') \qquad (\boldsymbol{a} \in P) . \tag{7.4}$$

We prove that this random linear function meets the requirements. It is enough to prove that (7.1) follows from the event (7.2, 7.3).

Obviously we have

$$\Theta(\boldsymbol{a}) \geq \underline{\Theta}(\boldsymbol{a}) \geq \underline{\Theta}(\boldsymbol{a}') + [\underline{\boldsymbol{\Psi}}(\boldsymbol{a}')]^T (\boldsymbol{a} - \boldsymbol{a}') . \tag{7.5}$$

From (7.3), we have

$$[\boldsymbol{\pi} - \underline{\boldsymbol{\Psi}}(\boldsymbol{a}')]^T (\boldsymbol{a} - \boldsymbol{a}') \leq \| \boldsymbol{\pi} - \underline{\boldsymbol{\Psi}}(\boldsymbol{a}') \| \, \| \boldsymbol{a} - \boldsymbol{a}' \|$$

$$\leq \sqrt{m\lfloor H \rfloor} \, \| \boldsymbol{\pi} - \underline{\boldsymbol{\Psi}}(\boldsymbol{a}') \|_{\max} \, \mathrm{Diam}(P) \leq \sqrt{m\lfloor H \rfloor} \, \tilde{\delta_2} \, \mathrm{Diam}(P) = \delta_2 .$$

It follows that

$$[\underline{\boldsymbol{\Psi}}(\boldsymbol{a}')]^T (\boldsymbol{a} - \boldsymbol{a}') \geq \boldsymbol{\pi}^T (\boldsymbol{a} - \boldsymbol{a}') - \delta_2 . \tag{7.6}$$

From (7.2), we also have $\underline{\Theta}(\boldsymbol{a}') \geq \underline{\zeta}$. Substituting this and (7.6) into the right-hand side of (7.5), we get $\Theta(\boldsymbol{a}) \geq \Lambda(\boldsymbol{a})$ which is the first part of (7.1). The second part follows from (7.2):

$$\Theta(\boldsymbol{a}') - \Lambda(\boldsymbol{a}') \;\leq\; \bar{\zeta} - (\underline{\zeta} - \delta_2) \;-\; \boldsymbol{\pi}^T(\,\boldsymbol{a}' \,-\, \boldsymbol{a}'\,) \;=\; (\bar{\zeta} - \underline{\zeta}) + \delta_2 \;\leq\; \delta_1 + \delta_2 \;=\; \delta\,.$$

Since the probablitity of the event (7.2, 7.3) is at least p, it follows that we have (7.1) with a probability at least p.

Well-generated Instances of the Random Linear Function. From a practical point of view, we work with realizations of the random linear function Λ. First we generate realizations of the random variables $\bar{\zeta}$ and $\underline{\zeta}$, and of the random vector $\boldsymbol{\pi}$. Then we construct a realization (instance) of Λ according to (7.4).

We say that the instance Λ of the random linear function is *well-generated*, if it was constructed from such instances $\bar{\zeta}$, $\underline{\zeta}$, and $\boldsymbol{\pi}$ that satisfy (7.2) and (7.3), respectively.

The construction of the random linear function is such that *the instance Λ is well-generated with a probability at least p.* Moreover, (7.1) follows from the well-generatedness of the instance Λ. I.e., using vector notations, it follows that

$$\Theta(\boldsymbol{a}) \;\geq\; \Lambda(\boldsymbol{a}) \quad (\boldsymbol{a} \in P) \qquad \text{and} \qquad \Theta(\boldsymbol{a}') - \Lambda(\boldsymbol{a}') \;\leq\; \delta\,.$$

Now let us observe another consequence of well-generatedness. Suppose that at $\boldsymbol{a}' \in P$, two random linear functions are constructed independently, according to the procedure described above.

– Given a tolerance $\delta > 0$ and a probability $1 > p > 0$, we construct Λ:

$$\Lambda(\boldsymbol{a}) \;=\; (\underline{\zeta} - \delta_2) \;+\; \boldsymbol{\pi}^T(\,\boldsymbol{a} \,-\, \boldsymbol{a}'\,) \qquad (\boldsymbol{a} \in P)\,.$$

– Given a tolerance $\delta^\star > 0$ and a probability $1 > p^\star > 0$, we construct $\Lambda^\star$:

$$\Lambda^\star(\boldsymbol{a}) \;=\; (\underline{\zeta^\star} - \delta_2^\star) \;+\; \boldsymbol{\pi}^{\star T}(\,\boldsymbol{a} \,-\, \boldsymbol{a}'\,) \qquad (\boldsymbol{a} \in P)\,.$$

We are going to prove the following statement: *If Λ and $\Lambda^\star$, respectively, are well-generated instances, then we have*

$$|\Lambda(\boldsymbol{a}) - \Lambda^\star(\boldsymbol{a})| \;\leq\; 2(\delta + \delta^\star) \qquad (\boldsymbol{a} \in P)\,.$$

The proof is simple: Well-generatedness of the instance Λ means that the instances $\bar{\zeta}$, $\underline{\zeta}$, and $\boldsymbol{\pi}$ satisfy (7.2) and (7.3), respectively.

Well-generatedness of the instance $\Lambda^\star$ means that the instances $\bar{\zeta^\star}$, $\underline{\zeta^\star}$, and $\boldsymbol{\pi}^\star$ satisfy the analogs of (7.2) and (7.3), respectively.

Thus we have

$$\begin{aligned} |\Lambda(\boldsymbol{a}) - \Lambda^\star(\boldsymbol{a})| \;&\leq\; |\underline{\zeta} - \underline{\zeta^\star}| \;+\; |\delta_2 - \delta_2^\star| \;+\; \|\boldsymbol{\pi} - \underline{\boldsymbol{\Psi}}(\boldsymbol{a}')\|\|\boldsymbol{a} - \boldsymbol{a}'\| + \|\underline{\boldsymbol{\Psi}}(\boldsymbol{a}') - \boldsymbol{\pi}^\star\|\|\boldsymbol{a} - \boldsymbol{a}'\| \\ &\leq\; 2(\delta + \delta^\star) \qquad\qquad (\boldsymbol{a} \in P)\,. \end{aligned}$$

Lipschitz Condition. According to the remark at the end of Section 5.3.2, we have

$$|\,\pi_{(h-1)m+k}\,| \;\leq\; q_k\,\mathrm{E}(|\phi_h{}'|) + \tilde{\delta_2} \qquad (k = 1,\ldots,m\,;\; h = 1,\ldots,\lfloor H \rfloor)$$

assumed that (7.3) holds. The probability of this event is at least p.

If the above inequalities should not hold, then let us trim the deviant components of $\boldsymbol{\pi}$ to fit into the appropriate intervals. (Let this trimming be part of the definition of the random vector $\boldsymbol{\pi}$.) Thus we have

$$\|\boldsymbol{\pi}\| \;\leq\; \sqrt{\sum_{k,\,h} \left(q_k\,\mathrm{E}(|\phi_h{}'|) \;+\; \frac{\delta}{\sqrt{m\lfloor H \rfloor}\,\mathrm{Diam}(P)} \right)^2}\,.$$

Let us introduce the notation $\mathcal{Q}(\delta)$ for the right-hand-side expression. The random linear function $\mathit{\Lambda}$ (constructed with the trimmed $\boldsymbol{\pi}$) satisfies the Lipschitz condition with the constant $\mathcal{Q}(\delta)$.

If the instance $\mathit{\Lambda}$ is well-generated, then no trimming takes place.

7.1 Tightening the Gap between the Objective Value and its Lower Bound

Given a tolerance $\delta_1 > 0$ and a probability $1 > p_1 > 0$, we need random variables $\bar{\zeta}$ and $\underline{\zeta}$ such that

$$\mathrm{P}\left(\bar{\zeta} \;\geq\; \Theta(A') \;\; \textit{and} \;\; \underline{\Theta}(A') \;\geq\; \underline{\zeta} \right) \;\geq\; p_1\,, \tag{7.7}$$

and $\bar{\zeta} - \underline{\zeta} \leq \delta_1$.

The selection of appropriate basis points for the approximation $\underline{\Theta}$ is part of the construction. Moreover, realizations of the random variables $\bar{\zeta}$ and $\underline{\zeta}$ must be easy to generate.

First let us estimate the objective value $\Theta(A')$ as described in Chapter 6. For this estimation, let us use a tolerance significantly smaller than δ_1, and a reliability parameter significantly higher than p_1.

Let us retain the basis points constructed in course of the above procedure. Let us consider the function $\underline{r}$ as defined on those basis points. Now let us estimate $\underline{\Theta}(A')$. That is the expectation of the expression

$$\boldsymbol{q}'^{T}\,[\,A'\,\boldsymbol{\phi}' \;-\; \boldsymbol{b}'\,]_- \;+\; \underline{r}\,(\,[\,A'\,\boldsymbol{\phi}' \;-\; \boldsymbol{b}'\,]_+\,)\,.$$

The first term is easy to compute. The second term can be estimated as

$$\frac{1}{M} \sum_{j=1}^{M} \underline{r}(\,[\boldsymbol{\eta}^{(j)}]_+\,)\,,$$

where $\boldsymbol{\eta}^{(j)}$ $(j = 1,\ldots,M)$ designate mutually independent random vectors that have the same distribution as $\boldsymbol{\eta} \;=\; A'\,\boldsymbol{\phi}' \;-\; \boldsymbol{b}'$.

If the gap $\Theta(A') - \underline{\Theta}(A')$ is too large, then new basis points must be added. It can be done in a way similar to the approximation scheme described in Section 6.2. Then we again estimate $\underline{\Theta}(A')$ as defined on the extended set of basis points.

This procedure can be iterated until the gap becomes tight enough.

7.2 Estimation of a Subgradient of the Approximate Objective Function

Given a set of basis points, a tolerance $\tilde{\delta}_2 > 0$ *and a probability* $1 > \tilde{p}_2 > 0$, *we need a random* $m\lfloor H \rfloor$*-vector* $\boldsymbol{\pi}$ *such that*

$$\mathrm{P}\left(\| \boldsymbol{\pi} - \underline{\boldsymbol{\Psi}}(A') \|_{\max} \leq \tilde{\delta}_2 \right) \geq \tilde{p}_2 \,, \tag{7.8}$$

where the approximation $\underline{\Theta}$ *is based on the given set of points, and the* $m\lfloor H \rfloor$*-vector* $\underline{\boldsymbol{\Psi}}(A')$ *is the subgradient of* $\underline{\Theta}$ *at* A'.

Moreover, realizations of the random vector $\boldsymbol{\pi}$ *must be easy to generate.*

In Section 5.3.2 we saw that the subgradient $\underline{\boldsymbol{\Psi}}(A')$ can be computed as the expectation of the $m\lfloor H \rfloor$-vector

$$\left[\boldsymbol{\mathcal{G}}(A'\boldsymbol{\phi}' - \boldsymbol{b}') \right]^T \left(\phi_1{}'I, \ \ldots, \ \phi_{\lfloor H \rfloor}{}'I \right) \,, \tag{7.9}$$

with respect to the random vector $\boldsymbol{\phi}'$.

The component $(h-1)m + k \quad (k = 1, \ldots, m \,;\ h = 1, \ldots, \lfloor H \rfloor)$ of the vector (7.9) is

$$\mathcal{G}_k(A'\boldsymbol{\phi}' - \boldsymbol{b}') \, \phi_h{}' \,.$$

The expectation of this component is

$$\mathrm{E}\left(\mathcal{G}_k(A'\boldsymbol{\phi}' - \boldsymbol{b}') \ \phi_h{}' \right) \ = \ \mathrm{E}\left(\mathcal{G}_k(\boldsymbol{\eta}) \ \mathrm{E}(\phi_h{}' \,|\, \boldsymbol{\eta}) \right) \,,$$

where the left-hand-side expectation is considered with respect to the random $\lfloor H \rfloor$-vector $\boldsymbol{\phi}'$, and the right-hand-side expectation with respect to the random m-vector $\boldsymbol{\eta} = A'\boldsymbol{\phi}' - \boldsymbol{b}'$. The right-hand-side expectation can be estimated by a realization of the random variable

$$\frac{1}{\tilde{M}} \sum_{j=1}^{\tilde{M}} \mathcal{G}_k\left(\tilde{\boldsymbol{\eta}}^{(j)}\right) \mathrm{E}\left(\phi_h{}' \,\middle|\, \tilde{\boldsymbol{\eta}}^{(j)} \right) \,, \tag{7.10}$$

where $\tilde{\boldsymbol{\eta}}^{(j)} \quad (j = 1, \ldots, \tilde{M})$ are mutually independent random vectors that have the same distribution as $\boldsymbol{\eta}$.

In order to generate a realization of the random variable (7.10), we need to generate realizations $\tilde{\boldsymbol{y}}^{(j)} \quad (j = 1, \ldots, \tilde{M})$ of the random vector $\boldsymbol{\eta}$. See Deák (1990) for the generation of normally distributed random vectors. Given A' and $\tilde{\boldsymbol{y}}^{(j)}$, the expectation $\mathrm{E}\left(\phi_h{}' \,\middle|\, A'\boldsymbol{\phi}' - \boldsymbol{b}' = \tilde{\boldsymbol{y}}^{(j)} \right)$ is easy to compute.

Let us define the random $m\lfloor H \rfloor$-vector $\boldsymbol{\pi}$ accordingly, as follows:

$$\pi_{(h-1)m+k} \;=\; \frac{1}{\tilde{M}} \sum_{j=1}^{\tilde{M}} \mathcal{G}_k\left(\tilde{\boldsymbol{\eta}}^{(j)}\right) \mathrm{E}\left(\phi_h{}' \,\middle|\, \tilde{\boldsymbol{\eta}}^{(j)}\right) \qquad (\, k = 1, \ldots, m \,;\; h = 1, \ldots, \lfloor H \rfloor \,) \,.$$

In Section 7.2.1 we compute the sample size $\tilde{M}$ which ensures that the probability requirement (7.8) is satisfied.

7.2.1 Computation of the Sample Size for the Estimation of a Subgradient

The size $\tilde{M}$ of the sample required to achieve a prescribed accuracy and reliability can be computed by using the multidimensional central limit theorem. Let us consider the random vector (7.9). We know that the expectation vector $\tilde{\boldsymbol{\mu}} = \underline{\Psi}(A')$ exists and is finite. First we prove that the covariance matrix is finite. Since we have $|\mathrm{Cov}(\,\xi,\,\zeta\,)| \leq \sqrt{\mathrm{Var}(\xi)\mathrm{Var}(\zeta)}$ for any random variables ξ and ζ, it is enough to find upper bounds for the diagonal elements. The component $(h-1)m+k \quad (k = 1, \ldots, m \,;\; h = 1, \ldots, \lfloor H \rfloor)$ of the vector (7.9) is

$$\mathcal{G}_k(\, A'\boldsymbol{\phi}' \,-\, \boldsymbol{b}' \,)\, \phi_h{}' \,.$$

Proposition 28 *We have*

$$\mathrm{Var}\,(\, \mathcal{G}_k(\, A'\boldsymbol{\phi}' \,-\, \boldsymbol{b}' \,)\, \phi_h{}' \,) \;\leq\; q_k^2\, \mathrm{E}\left(\phi_h{}'^2\right) \,.$$

(The second moment on the right-hand side is known and finite.)

Proof. Obviously we have $\mathrm{Var}(\xi) \leq \mathrm{E}(\xi^2)$ for any random variable ξ. It follows that

$$\mathrm{Var}\,(\, \mathcal{G}_k(\, A'\boldsymbol{\phi}' \,-\, \boldsymbol{b}' \,)\, \phi_h{}' \,) \;\leq\; \mathrm{E}\left[(\, \mathcal{G}_k(\, A'\boldsymbol{\phi}' \,-\, \boldsymbol{b}' \,)\, \phi_h{}' \,)^2\right] \,.$$

From the construction of the subgradient $\mathcal{G}$, we have $\left|\mathcal{G}_k(\boldsymbol{\eta})\right| \leq q_k$. □

We assume that the components of the random vector (7.9) are linearly independent. (Otherwise let us choose a maximal linearly independent subset of the components. The following arguments will hold.) Let $\tilde{C}$ designate the (positive definite) covariance matrix.

According to a multidimensional central limit theorem, the distribution of the random vector $\boldsymbol{\pi}$ can be approximated by the $m\lfloor H \rfloor$-variate $N(\tilde{\boldsymbol{\mu}}, \frac{1}{\tilde{M}}\tilde{C})$-distribution. (I.e., the normal distribution having expectation vector $\tilde{\boldsymbol{\mu}} = \underline{\boldsymbol{\Psi}}(A')$ and covariance matrix $\frac{1}{\tilde{M}}\tilde{C}$.) Proof of that theorem can be found in Section 4.2.3 of Anderson (1958) or in Section 9.2 of Wilks (1962). Hence the random vector $\boldsymbol{\pi} - \underline{\boldsymbol{\Psi}}(A')$ has the $m\lfloor H \rfloor$-variate $N(\mathbf{0}, \frac{1}{\tilde{M}}\tilde{C})$-distribution.

Šidak (1968) proved the following theorem: *If* $(X_1, \ldots, X_\ell)$ *has a normal distribution with mean* $\mathbf{0}$ *and an arbitrary correlation matrix, then*

$$\mathrm{P}(\, |X_1| < c_1,\, \ldots,\, |X_\ell| < c_\ell \,) \;\geq\; \prod_{i=1}^{\ell} \mathrm{P}(\, |X_i| < c_i \,)$$

In the present case, we have $\ell = m\lfloor H \rfloor$, $c_1 = \ldots = c_\ell = \tilde{\delta}_2$; the variable $X_{(h-1)m+k}$ has variance $\frac{1}{\tilde{M}} \operatorname{Var}\left(\mathcal{G}_k(\boldsymbol{\eta}) \operatorname{E}(\phi_h{}' \,|\, \boldsymbol{\eta}) \right)$ $(k = 1, \ldots, m, \;\; h = 1, \ldots, \lfloor H \rfloor)$.

Let us introduce the notation

$$\tilde{\sigma}_{(h,k)} \;=\; \sqrt{\operatorname{Var}\left(\mathcal{G}_k(\boldsymbol{\eta}) \operatorname{E}(\phi_h{}' \,|\, \boldsymbol{\eta}) \right)} \qquad (k = 1, \ldots, m, \;\; h = 1, \ldots, \lfloor H \rfloor) .$$

As a consequence of Šidak's theorem, we have

$$\operatorname{P}\Big(\| \boldsymbol{\pi} \,-\, \underline{\boldsymbol{\Psi}}(A') \|_{\max} \;\le\; \tilde{\delta}_2 \Big) \tag{7.11}$$

$$\geq \prod_{\substack{k = 1, \ldots, m \\ h = 1, \ldots, \lfloor H \rfloor}} \int_{-\tilde{\delta}_2}^{\tilde{\delta}_2} \frac{\sqrt{\tilde{M}}}{\sqrt{2\pi}\, \tilde{\sigma}_{(h,k)}} \exp\left(- \frac{\tilde{M} x^2}{2\, \tilde{\sigma}^2_{(h,k)}} \right) dx \,.$$

The term (k, h) of the above product can be tarnsformed as follows:

$$\int_{-\tilde{\delta}_2 \sqrt{\tilde{M}} / \tilde{\sigma}_{(h,k)}}^{-\tilde{\delta}_2 \sqrt{\tilde{M}} / \tilde{\sigma}_{(h,k)}} \frac{1}{\sqrt{2\pi}} \; e^{-\frac{1}{2} y^2} \, dy \;=\; 2\,\Phi\left(\frac{\tilde{\delta}_2 \sqrt{\tilde{M}}}{\tilde{\sigma}_{(h,k)}} \right) - 1 \,,$$

where Φ is the distribution function of the $N(0,1)$ -distribution. Let us introduce the notation

$$\tilde{s}_{(h,k)} \;=\; q_k \sqrt{\operatorname{E}\left(\phi_h{}'^2 \right)} \qquad (k = 1, \ldots, m, \;\; h = 1, \ldots, \lfloor H \rfloor) .$$

From Proposition 28, and the monotonicity of Φ, it follows that the following expression is a lower bound for the probability (7.11)

$$\prod_{\substack{k = 1, \ldots, m \\ h = 1, \ldots, \lfloor H \rfloor}} \left(2\,\Phi\left(\frac{\tilde{\delta}_2 \sqrt{\tilde{M}}}{\tilde{s}_{(h,k)}} \right) - 1 \right) . \tag{7.12}$$

Let us select the sample size $\tilde{M}$ such that the above expression be at least $\tilde{p}_2$.

Remark. In a realistic case, we have $H = 75$, $m = 3$. The large number of multipliers in the product (7.12) causes no unmanageable problem. Assume that with a moderate sample size $\tilde{M}$, we can achieve that each multiplier of the product (7.12) is at least 0.9545. This means

$$\frac{\tilde{\delta}_2 \sqrt{\tilde{M}}}{\tilde{s}_{(h,k)}} \;\ge\; 2 \qquad (k = 1, \ldots, m, \quad h = 1, \ldots, \lfloor H \rfloor) .$$

By quadrupling the sample size, we can achieve

$$\frac{\tilde{\delta}_2 \sqrt{\tilde{M}}}{\tilde{s}_{(h,k)}} \;\ge\; 4 \,, \quad \text{and thus} \quad 2\,\Phi\left(\frac{\tilde{\delta}_2 \sqrt{\tilde{M}}}{\tilde{s}_{(h,k)}} \right) - 1 \;\ge\; 0.9999 \quad \text{for each } h, k.$$

Hence the probability (7.12) grows to at least $0.9999^{225} \geq 0.985$.

Further, if we make the sample size 7 times as big as the original, then the probability (7.12) grows to about 0.99997. Assuming that we set $p_1 = p_2 = \sqrt{p}$, and the objective value can be estimated with similar reliability, the reliability parameter of the whole process may reach $p = 0.99994$.

Remark. The expression $\tilde{s}^2_{(h,k)}$ is a rather raw upper estimate of the variance $\tilde{\sigma}^2_{(h,k)}$.

Better results are expected from multi-stage tests, where the variances $\tilde{\sigma}^2_{(h,k)}$ $(k = 1, \ldots, m, \quad h = 1, \ldots, \lfloor H \rfloor)$ are estimated using sample variances of $m\lfloor H \rfloor$ mutually independent samples.

Chapter 8

Solution Method for the Dynamic Problem

The aim is to minimize a convex function that can be estimated by simulation. Given a point of the feasible domain, the objective value and subgradient can be estimated with prescribed accuracy and reliability. The more accurate or reliable estimates are required, the bigger the computational effort.

Since the estimation of the objective value and subgradient requires a relatively big effort, the estimate data are worth retaining and reusing. A feasible direction method discards these data once it moved to a new point. The classical cutting-plane method of Kelley (1960) does accumulate all the information obtained, but is still known as very slow. (Slowness is the consequence of instability.) Bundle methods are a refinement of the cutting-plane method. (A comprehensive treatment of the bundle methods can be found in Hiriart-Urruty and Lemaréchal (1993).) The idea is to maintain a stability center or prox-center, that is, to distinguish one of the iterates generated that far. The prox-center is updated every time a significantly better point was found. Roaming away from the current prox-center is penalized. Thus bundle methods reduce the influence of the inaccuracy of the cutting-plane approximation, thereby reducing instability. More recently, Lemaréchal, Nemirovskii, and Nesterov (1995) proposed such variants of the bundle methods, that the prox-center can be forgotten about. (Though it exists implicitly, and attracts the path along which new iterates are visited.)

We develop a stochastic version of the Level Method of Lemaréchal, Nemirovskii, and Nesterov. At any stage, good upper and lower bounds for the optimal objective value are available. We are going to exploit this feature. At the beginning, only rough estimates of the objective value and gradient are needed. As the optimum is gradually approached, more and more accurate data are computed. Such heuristics are already employed in the code of Szántai (1988). The present procedure has the advantage that the required accuracy is known at each step, hence better coordination of efforts is possible.

In Section 8.1 the Level Method is briefly described. In Section 8.2 we develop a stochastic version of the Level Method.

8.1 Description of the Level Method

Consider the problem of minimizing a convex fiunction f over a convex set $Q \subset \mathbb{R}^K$. Having generated the iterates $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i \in Q$, and using an oracle to compute the function values $f(\boldsymbol{x})$ and the subgradients $\nabla f(\boldsymbol{x})$, the supporting hyperplane approximation of f is

$$f_i(\boldsymbol{x}) = \max_{1 \le j \le i} \left\{ f(\boldsymbol{x}_j) + [\nabla f(\boldsymbol{x}_j)]^T (\boldsymbol{x} - \boldsymbol{x}_j) \right\} \qquad (\boldsymbol{x} \in Q) .$$

Clearly, f_i is a piecewise linear, convex function that underestimates f. Moreover, we have $f_i(\boldsymbol{x}_j) = f(\boldsymbol{x}_j)$ $(j \le i)$ and $f_j \le f_i$ $(j \le i)$.

The aim is to enclose the optimal solution in a contracting sequence of sets. If the iterates $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i, \ldots$ are properly selected, then the functions f_i are more and more accurate approximations in a contracting neighborhood of the optimum.

The classical cutting plane method computes the iterate $\boldsymbol{x}_{i+1}$ by minimizing f_i over Q. The problem is that the inaccuracy of the approximation may have an excessive effect on the selection of new iterates.

The Level Method of Lemaréchal, Nemirovskii, and Nesterov (1995) is a natural refinement of the classical cutting plane method. We describe the method briefly. The best function value obtained until the ith step,

$$U_i = \min_{1 \le j \le i} f(\boldsymbol{x}_j) ,$$

will be used as an upper bound for the optimum. A lower bound is the minimum of the ith approximating function:

$$L_i = \min_{\boldsymbol{x} \in Q} f_i(\boldsymbol{x}) .$$

The gap between the above bounds is $\Delta_i = U_i - L_i$.

Clearly, the sequence of the upper bounds U_i is monotone decreasing, and the sequence of the lower bounds L_i is monotone increasing. Hence the gap is tightening at each step.

Let $0 < \lambda < 1$ be some preset parameter. Consider the level set

$$Q_i = \{ \boldsymbol{x} \in Q \mid f_i(\boldsymbol{x}) \le L_i + \lambda \Delta_i \} .$$

The next iterate $\boldsymbol{x}_{i+1}$ is computed by projecting $\boldsymbol{x}_i$ to the level set Q_i. That is,

$$\boldsymbol{x}_{i+1} = \arg \min_{\boldsymbol{x} \in Q_i} \operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{x}),$$

where dist means the Eucledian distance. (Setting $\lambda = 0$ gives the classical method. With non-extremal setting, the level sets stabilize the procedure.)

Assume that the feasible region is bounded, and has the diameter D. Assume that f is Lipschitz-continuous with the constant κ. We need an upper bound for the number of iterations required to decrease the gap under a prescribed tolerance. The key argument is the following

Proposition 29 *Consider a sequence of iterations at the end of which the gap has not been reduced much. Namely, let t and s be natural numbers, $t < s$, and assume that the following inequality holds*

$$(1-\lambda)\Delta_t \le \Delta_s. \tag{8.1}$$

Then the number of iterations performed while getting from $\boldsymbol{x}_t$ to $\boldsymbol{x}_s$, can not be greater than

$$\left(\frac{\kappa D}{(1-\lambda)\Delta_s}\right)^2.$$

Proof. Consider the intervals

$$[\,L_t,\,U_t\,] \supseteq [\,L_s,\,U_s\,],$$

the lengths of which are Δ_t and Δ_s, respectively. The point $L_t + \lambda\Delta_t$ divides the first interval into two subintervals, the upper one having a length of $(1-\lambda)\Delta_t$. According to the assumption (8.1), the upper subinterval can not contain $[L_s, U_s]$ in the interior. Hence we have $L_s \le L_t + \lambda\Delta_t$. Using the notation

$$\boldsymbol{u}_s = \arg\min_{\boldsymbol{x}\in Q} f_s(\boldsymbol{x}),$$

we just proved $\boldsymbol{u}_s \in Q_t$. Similarly, each level set Q_i $(t \le i \le s)$ contains $\boldsymbol{u}_s$.

The point $\boldsymbol{u}_s$ can be viewed as a stability center, though it does not lie on the path of the method. (That is, $\boldsymbol{u}_s$ is not among the iterates $\boldsymbol{x}_i$.) We are going to show that $\boldsymbol{u}_s$ indeed attracts the path along which new iterates are visited. By definition, the iterate $\boldsymbol{x}_{i+1}$ is the projection of $\boldsymbol{x}_i$ to the level set Q_i. Hence the iterates are getting closer and closer to the point $\boldsymbol{u}_s$. Namely, from the properties of the projection, we have

$$\operatorname{dist}(\boldsymbol{x}_{i+1}, \boldsymbol{u}_s)^2 \;\le\; \operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{u}_s)^2 \;-\; \operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{x}_{i+1})^2 \qquad (t \le i \le s-1). \tag{8.2}$$

We are going to find a lower bound for the decrease $\operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{x}_{i+1})^2$. Obviously we have

$$f_i(\boldsymbol{x}_i) = f(\boldsymbol{x}_i) \;\ge\; U_i \quad \text{and} \quad f_i(\boldsymbol{x}_{i+1}) \;\le\; L_i + \lambda\Delta_i \qquad (t \le i \le s-1).$$

From the above inequalities, we get

$$f_i(\boldsymbol{x}_i) - f_i(\boldsymbol{x}_{i+1}) \;\ge\; (U_i - L_i) - \lambda\Delta_i \;=\; (1-\lambda)\Delta_i \qquad (t \le i \le s-1).$$

Since the function f_i inherits Lipschitz-continuity from f, it follows that

$$\operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{x}_{i+1}) \;\ge\; \frac{1}{\kappa}(1-\lambda)\Delta_i.$$

By substituting this into the inequality (8.2), we get

$$\operatorname{dist}(\boldsymbol{x}_{i+1}, \boldsymbol{u}_s)^2 \;\le\; \operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{u}_s)^2 - \left(\frac{1}{\kappa}(1-\lambda)\Delta_i\right)^2 \qquad (t \le i \le s-1).$$

We have $\Delta_i \geq \Delta_s$, hence we may substitute Δ_s for Δ_i in the right-hand side of the above expression. Moreover, D being the diameter of the feasible domain, we have

$$\mathrm{dist}(\boldsymbol{x}_t, \boldsymbol{u}_s) \leq D\,.$$

Hence the number of steps performed while getting from $\boldsymbol{x}_t$ to $\boldsymbol{x}_s$, can not be greater then

$$\frac{D^2}{\left(\frac{1}{\kappa}\,(1-\lambda)\Delta_s\right)^2}\,.$$

□

Corollary 30 *Consider a sequence of iterations at the end of which the gap has not been reduced much. Namely, let $t < s$, and assume that the following inequality holds*

$$(1-\lambda)\Delta_t \leq \Delta_s\,.$$

Then the number of iterations performed while getting from $\boldsymbol{x}_t$ to $\boldsymbol{x}_s$, can not be greater than

$$\left(\frac{\kappa D}{(1-\lambda)^2\Delta_t}\right)^2\,.$$

Theorem 31 *To obtain a gap smaller then ϵ, it suffices to perform*

$$c\,\left(\frac{\kappa D}{\epsilon}\right)^2$$

iterations. (c is a constant that depends only on λ.)

Proof. Starting from $\boldsymbol{x}_1$, consider the maximal sequence of iterations :

$$\boldsymbol{x}_1 \to \boldsymbol{x}_2\,,\ \ldots,\ \boldsymbol{x}_{s-1} \to \boldsymbol{x}_s\,,$$

at the end of which the gap has not been reduced much. That is, the following inequalities hold:

$$(1-\lambda)\Delta_1 \leq \Delta_s\,, \quad (1-\lambda)\Delta_1 > \Delta_{s+1}\,.$$

The iteration $\boldsymbol{x}_s \to \boldsymbol{x}_{s+1}$ will be called *critical*.

The above construction is repeated starting from $\boldsymbol{x}_{s+1}$. Thus the iterations are grouped into sequences, and the sequences are separated with critical iterations. Assume there are k sequences. Denote by $\Delta^{(\ell)}$ $(1 \leq \ell \leq k)$ the gap at the start of the ℓth sequence. (E.g., we have $\Delta^{(1)} = \Delta_1$, $\Delta^{(2)} = \Delta_{s+1}$.) From the definitions, it follows that

$$(1-\lambda)\Delta^{(\ell)} \ > \ \Delta^{(\ell+1)} \qquad (\ell = 1, \ldots, k-1)\,.$$

We have $\Delta^{(k)} > \epsilon$. (Otherwise the process stops before the kth sequence.) It follows that

$$(1-\lambda)^{k-\ell}\,\Delta^{(\ell)} \ > \ \epsilon \qquad (\ell = 1, \ldots, k)\,. \tag{8.3}$$

Hence from Corollary 30, it follows that the length of the ℓth sequence is not greater than

$$\left(\frac{\kappa D}{(1-\lambda)^2\,\epsilon}\right)^2 (1-\lambda)^{2(k-\ell)} \qquad (\ell = 1, \ldots, k)\,.$$

The number of non-critical iterations can be computed by summing up the lengths of the sequences. An upper bound is

$$\begin{aligned}\left(\tfrac{\kappa D}{(1-\lambda)^2\,\epsilon}\right)^2 \sum_{\ell=1}^{k} (1-\lambda)^{2(k-\ell)} &\le \left(\tfrac{\kappa D}{(1-\lambda)^2\,\epsilon}\right)^2 \sum_{(k-\ell)=0}^{\infty} (1-\lambda)^{2(k-\ell)} \\ &\le \left(\tfrac{\kappa D}{(1-\lambda)^2\,\epsilon}\right)^2 \tfrac{1}{1-(1-\lambda)^2} \;=\; \left(\tfrac{\kappa D}{\epsilon}\right)^2 \tfrac{1}{\lambda(1-\lambda)^4(2-\lambda)}\,.\end{aligned}$$

The number of the critical iterations, k, is negligable compared to the above expression: from (8.3), it follows that

$$(1-\lambda)^{k-1}\,\kappa D > \epsilon\,,$$

which gives the following upper bound for the number of the critical iterations:

$$C\,\log\left(\frac{\kappa D}{\epsilon}\right) + 1,$$

where C depends only on λ. (If we set $\lambda = \frac{1}{2}$, then we get $C = 1$.) □
(In Lemaréchal, Nemirovskii, and Nesterov (1995), a somewhat sharper bound is proved. The construction of the sequences is done backwards, starting at the last step. Accordingly, Proposition 29 is used, and not Corollary 30. The modification was made for the sake of simplicity.)

Remark. A drawback of the method is the constantly increasing size of the linear and quadratic problems that need to be solved at each iteration. In order to overcome this, reasonable strategies are wanted to discard some of the cutting planes, or aggregate the information obtained.

Lemaréchal, Nemirovskii, and Nesterov (1995) report on successful applications of the Level Method to a variety of problems. (The largest of which had 442 variables.) The following reduction strategy is applied: After each critical iteration, all the non-binding cutting planes are eliminated. Namely, if the sth iteration turns out to be critical, then they eliminate all the cutting planes which are inactive at the determination of the gap Δ_{s+1}. It is easily seen that the theoretical efficiency estimates are preserved by the above reduction strategy.

Concerning practical behavior, the numbers of the iterations are reported to have increased by at most 15%. (The constant $\lambda = \frac{1}{2}$ was used in each test run.) Moreover, the cardinality of the *bundle* (i.e., the number of the cutting planes stored simultaneously) never exceeded $2K$. This implies that the number of steps between two consecutive critical iterations may be much smaller in practice than the theoretical bound in Proposition 29.
□

8.2 A Stochastic Version of the Level Method

We are going to adapt the Level Method to the stochastic problem of minimizing $f = \Theta$ over $Q = P$.

In order to implement the Level Method we should be able to generate a supporting hyperplane for the objective function at an iterate $\boldsymbol{x}_i$. We can not do that, but we have a good substitute: *Given a tolerance $\delta_i > 0$ and a probability $1 > p > 0$, we can construct a random linear function Λ_i such that*

$$f(\boldsymbol{x}) \geq \Lambda_i(\boldsymbol{x}) \quad (\boldsymbol{x} \in Q) \qquad \textit{and} \qquad f(\boldsymbol{x}_i) - \Lambda_i(\boldsymbol{x}_i) \leq \delta_i \tag{8.4}$$

holds simultaneously with a probability at least p. (The accuracy δ_{i+1} to be prescribed for the $(i+1)$th step will be computed with regard to the gap Δ_i.)

Well-generated Instances of the Random Linear Function. We actually need a somewhat stronger notion: *well-generatedness*. It applies to instances (realizations) of the random linear function. It means two things:

1: Suppose that given a tolerance $\delta_i > 0$ and a probability $1 > p > 0$, we construct the random linear function Λ_i. If the instance Λ_i of this random function is well-generated, then (8.4) holds.

2: Suppose that at the iterate $\boldsymbol{x}_i \in Q$, two random linear functions are constructed independently.

- Given a tolerance $\delta_i > 0$ and a probability $1 > p > 0$, we construct Λ_i.
- Given a tolerance $\delta_i^\star > 0$ and a probability $1 > p^\star > 0$, we construct $\Lambda_i^\star$.

If Λ_i and $\Lambda_i^\star$, respectively, are well-generated instances, then we have

$$|\Lambda_i(\boldsymbol{x}) - \Lambda_i^\star(\boldsymbol{x})| \leq 2(\delta_i + \delta_i^\star) \qquad (\boldsymbol{x} \in Q)\,.$$

The construction of the random linear function is such that *if we use the parameters δ_i and p, then the instance Λ_i is well-generated with a probability at least p.*

Lipschitz Condition. The construction of the random linear function has another useful feature: *if we use the parameters δ_i and p, then the instance Λ_i satisfies the Lipschitz condition with a constant that depends only on δ_i.* (The smaller the tolerance, the smaller the Lipschitz constant $\mathcal{Q}(\delta_i)$.)

In Section 8.2.1 we consider the case when the supporting hyperplanes can not be computed exactly, but the inaccuracy remains under a prescribed tolerance. In Section 8.2.2, we describe a way of handling random data.

8.2.1 Handling Inexact Data

In this section we assume that every cutting plane is well-generated, and hence the event (8.4) always occurs. I.e., the data are inexact, but the inaccuracy always remains under a prescribed tolerance.

The ith approximating function will be defined as

$$f_i(\boldsymbol{x}) = \max_{1 \le j \le i} \Lambda_j(\boldsymbol{x}) \qquad (\boldsymbol{x} \in Q)\,.$$

Clearly, f_i is a piecewise linear, convex function that underestimates f. Moreover, we have $f_i(\boldsymbol{x}_j) \ge \Lambda_j(\boldsymbol{x}_j) \ge f(\boldsymbol{x}_j) - \delta_j \ \ (j \le i)$, and $f_j \le f_i \ \ (j \le i)$.

At the ith step we have the following upper bound for the optimum:

$$U_i = \min_{1 \le j \le i} \left(\Lambda_j(\boldsymbol{x}_j) + \delta_j \right).$$

A lower bound is the minimum of the ith approximating function:

$$L_i = \min_{\boldsymbol{x} \in Q} f_i(\boldsymbol{x})\,.$$

Clearly, the sequence of the upper bounds U_i is monotone decreasing, and the sequence of the lower bounds L_i is monotone increasing. Hence the gap $\Delta_i = U_i - L_i$ is tightening at each step.

Let $0 < \lambda < 1$ be some preset parameter. Consider the level set

$$Q_i = \{\, \boldsymbol{x} \in Q \mid f_i(\boldsymbol{x}) \le L_i + \lambda \Delta_i \,\}\,.$$

The next iterate $\boldsymbol{x}_{i+1}$ is computed by projecting $\boldsymbol{x}_i$ to the level set Q_i. That is,

$$\boldsymbol{x}_{i+1} = \arg\min_{\boldsymbol{x} \in Q_i} \operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{x}),$$

Let us select a starting accuracy $\delta_1 > 0$ arbitrarily. The accuracy δ_{i+1} will depend on $\Delta_i \ \ (i = 1, 2, \ldots)$, as follows. Let us select a constant parameter $0 < \gamma < (1-\lambda)^2$, and define

$$\delta_{i+1} = \gamma \Delta_i \quad (i = 1, 2, \ldots)\,.$$

Clearly, the sequence $\delta_{i+1} \ \ (i = 1, 2, \ldots)$ is monotone decreasing. Hence we have an upper bound $\bar{\delta}$ for the tolerances. It follows that the functions $f_i \ \ (i = 1, 2, \ldots)$ are Lipschitz-continuous with the constant $\kappa = \mathcal{Q}(\bar{\delta})$.

The following proposition is the analog of Proposition 29

Proposition 32 *Consider a sequence of iterations at the end of which the gap has not been reduced much. Namely, let $t < s$, and assume that the following inequality holds*

$$(1 - \lambda)\Delta_t \le \Delta_s\,. \tag{8.5}$$

Then the number of iterations performed while getting from $\boldsymbol{x}_t$ to $\boldsymbol{x}_s$, can not be greater than

$$1 + \left(\frac{\kappa D}{\left(1 - \lambda - \frac{\gamma}{1-\lambda}\right) \Delta_s} \right)^2 .$$

Proof. We will count the number of iterations while getting from $\boldsymbol{x}_{t+1}$ to $\boldsymbol{x}_s$. (The 1 in the above expression is for the first iteration, $\boldsymbol{x}_t \to \boldsymbol{x}_{t+1}$.)

Consider the intervals

$$[\, L_t, U_t \,] \supseteq [\, L_s, U_s \,] ,$$

the lengths of which are Δ_t and Δ_s, respectively. The point $L_t + \lambda\Delta_t$ divides the first interval into two subintervals, the upper one having a length of $(1-\lambda)\Delta_t$. According to the assumption (8.5), the upper subinterval can not contain $[L_s, U_s]$ in the interior. Hence we have $L_s \le L_t + \lambda\Delta_t$. Using the notation

$$\boldsymbol{u}_s = \arg\min_{\boldsymbol{x} \in Q} \; f_s(\boldsymbol{x}) ,$$

we just proved $\boldsymbol{u}_s \in Q_t$. Similarly, each level set Q_i $(t \le i \le s)$ contains $\boldsymbol{u}_s$.

By definition, the iterate $\boldsymbol{x}_{i+1}$ is the projection of $\boldsymbol{x}_i$ to the level set Q_i. Hence the iterates are getting closer and closer to the point $\boldsymbol{u}_s$. Namely, from the properties of the projection, we have

$$\operatorname{dist}(\boldsymbol{x}_{i+1}, \boldsymbol{u}_s)^2 \;\le\; \operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{u}_s)^2 \;-\; \operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{x}_{i+1})^2 \qquad (t \le i \le s-1) . \tag{8.6}$$

We are going to find a lower bound for the decrease $\operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{x}_{i+1})^2$. We have

$$f_i(\boldsymbol{x}_i) \ge \Lambda_i(\boldsymbol{x}_i) \ge U_i - \delta_i \quad \text{and} \quad f_i(\boldsymbol{x}_{i+1}) \le L_i + \lambda\Delta_i \qquad (t+1 \le i \le s-1) .$$

From the above inequalities and the definition of δ_i, we get

$$f_i(\boldsymbol{x}_i) - f_i(\boldsymbol{x}_{i+1}) \;\ge\; (U_i - L_i) - \lambda\Delta_i - \delta_i \;=\; (1-\lambda)\Delta_i - \gamma\Delta_{i-1} \qquad (t+1 \le i \le s-1) .$$

Here we have $(1-\lambda)\Delta_{i-1} \le \Delta_i$ as a consequence of the assumption (8.5). (That's why we need $t+1 \le i$.) Hence the above inequality can be continued as

$$f_i(\boldsymbol{x}_i) - f_i(\boldsymbol{x}_{i+1}) \;\ge\; (1-\lambda)\Delta_i - \gamma\Delta_{i-1} \;\ge\; \left(1 - \lambda - \frac{\gamma}{1-\lambda}\right) \Delta_i \qquad (t+1 \le i \le s-1) .$$

Since the function f_i is Lipschitz-continuous with the constant κ, it follows that

$$\operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{x}_{i+1}) \;\ge\; \frac{1}{\kappa} \left(1 - \lambda - \frac{\gamma}{1-\lambda}\right) \Delta_i$$

Substituting this into the inequality (8.6), we get

$$\operatorname{dist}(\boldsymbol{x}_{i+1}, \boldsymbol{u}_s)^2 \;\le\; \operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{u}_s)^2 - \left[\frac{1}{\kappa} \left(1 - \lambda - \frac{\gamma}{1-\lambda}\right) \Delta_i \right]^2 \qquad (t+1 \le i \le s-1) .$$

We have $\Delta_i \geq \Delta_s$, hence we may substitute Δ_s for Δ_i in the right-hand side of the above expression. Moreover, D being the diameter of the feasible set, we have

$$\mathrm{dist}(\boldsymbol{x}_{t+1}, \boldsymbol{u}_s) \leq D\,.$$

Hence the number of steps performed while getting from $\boldsymbol{x}_{t+1}$ to $\boldsymbol{x}_s$, can not be greater then

$$\frac{D^2}{\left[\,\frac{1}{\kappa}\left(1-\lambda-\frac{\gamma}{1-\lambda}\right)\Delta_s\,\right]^2}\,.$$

(The denominator is positive due to the selection $\gamma < (1-\lambda)^2$.) □

Corollary 33 *Consider a sequence of iterations at the end of which the gap has not been reduced much. Namely, let $t < s$, and assume that the following inequality holds*

$$(1-\lambda)\Delta_t \leq \Delta_s\,.$$

Then the number of iterations performed while getting from $\boldsymbol{x}_t$ to $\boldsymbol{x}_s$, can not be greater than

$$1 \;+\; \left(\frac{\kappa D}{[\,(1-\lambda)^2-\gamma\,]\;\Delta_t}\right)^2\,. \tag{8.7}$$

Theorem 34 *To obtain a gap smaller then ϵ, it suffices to perform*

$$\tilde{c}\;\left(\frac{\kappa D}{\epsilon}\right)^2$$

iterations. ($\tilde{c}$ is a constant that depends only on λ and on γ.)

The number of the critical iterations is at most

$$I^{\#} \;=\; \tilde{C}\,\log\left(\frac{\kappa D}{\epsilon}\right)\,+1.$$

($\tilde{C}$ depends only on λ.)

Moreover, the number of iterations between two subsequent critical iterations is at most

$$I \;=\; 1 \;+\; \left(\frac{\kappa D}{[\,(1-\lambda)^2-\gamma\,]\;\epsilon}\right)^2\,.$$

Proof. The proof is the same as that of Theorem 31, only Corollary 33 must be used instead of Corollary 30. (The number of the critical iterations is negligable compared to the above upper bound. Hence the 1 in the expression (8.7) does not alter the proof.)

We have

$$\tilde{c} \;=\; 1 \;\Big/\; \lambda\left[\,(1-\lambda)^2-\gamma\,\right]^2(2-\lambda)\,, \qquad \tilde{C} \;=\; 1 \;\Big/\; -\log(1-\lambda)\,.$$

□

Remark. According to the reports in Lemaréchal, Nemirovskii, and Nesterov (1995), the number of iterations between two subsequent critical iterations is at most $I = K$.

8.2.2 Handling Random Data

The cutting planes are well-generated only with a certain probability. (The inequalities (8.4) hold only with a certain probability.)

We are going to use two different levels of reliability, p and $p^\star$, in the construction of cutting planes. ($0 < p < p^\star < 1$.) Normally, a cutting plane is constructed with reliability p. After critical iterations, though, inactive cutting planes are eliminated, and the active cutting planes are reconstructed with the higher reliability $p^\star$. The set of the reconstructed cutting planes may be accepted as safe, to which we can recourse if inconsistency occurs in later iterations.

First we describe the procedure by words and formulas. Then we sketch the framework of the procedure. Finially, we provide formal proofs.

Description of the Procedure

Let us define the *bundle* as the set of the cutting planes stored simultaneously at a certain point of time. Every time a new cutting plane is constructed, we add it to the bundle. Let $\mathcal{B}_i \subset \{1, \ldots, i\}$ designate the set of the bundle subscripts at the i iteration. I.e., $j \in \mathcal{B}_i$ means that the cutting plane Λ_j belongs to the bundle at the ith iteration.

Normally, a cutting plane is constructed with reliability p. After critical iterations, though, inactive cutting planes are eliminated, and the active cutting planes are reconstructed with the higher reliability $p^\star$. The reduced and reconstructed bundle may be accepted as *safe*. (To be accepted, the reconstructed bundle must even pass an efficiency check that we discuss later: (8.13).) Let $\mathcal{B}_s^\#$ designate the sth accepted safe bundle; and $\Delta_s^\#$ the sth safe upper estimate on the gap. ($\Delta_s^\#$ is computed using the cutting planes of the sth safe bundle.)

Let us check the consistency of the bundle after each cutting plane generation. We must have

$$\Lambda_j(\boldsymbol{x}_j) + \delta_j \;\geq\; \Lambda_k(\boldsymbol{x}_j) \qquad (\, j, k \in \mathcal{B}_i \,)\,. \tag{8.8}$$

If the above check fails, then we are going to recourse to the latest safe bundle. If no inconsistency is detected, then we proceed as follows.

Let us define the ith cutting plane approximation as

$$f_i(\boldsymbol{x}) = \max_{j \in \mathcal{B}_i} \Lambda_j(\boldsymbol{x}) \qquad (\boldsymbol{x} \in Q)\,. \tag{8.9}$$

Accordingly, we compute

$$U_i = \min_{j \in \mathcal{B}_i} (\, \Lambda_j(\boldsymbol{x}_j) + \delta_j \,)\,, \qquad L_i = \min_{\boldsymbol{x} \in Q} f_i(\boldsymbol{x})\,, \qquad \Delta_i = U_i - L_i\,. \tag{8.10}$$

Let $\boldsymbol{x}_{i+1}$ be the projection of $\boldsymbol{x}_i$ to the level set $\{\, \boldsymbol{x} \in Q \mid f_i(\boldsymbol{x}) \leq L_i + \lambda\Delta_i \,\}$. Let us construct a new cutting plane at $\boldsymbol{x}_{i+1}$ with accuracy δ_{i+1} and reliability p, and add the new cutting plane to the bundle. (The prescribed accuracy, δ_{i+1} is computed from the latest safe upper estimate on the gap.)

The above procedure is repeated until we perceive a substantial decrease in the gap. I.e., until $\Delta_i < (1-\lambda)\Delta_s^{\#}$ holds, where $\Delta_s^{\#}$ is the latest safe upper estimate on the gap.

If we have $\Delta_i < (1-\lambda)\Delta_s^{\#}$ for some i, then the $(i-1)$th iteration is considered critical. (We will prove that such an iteration is reached with a positive probability.) After a critical iteration, we eliminate all the cutting planes which are inactive at the determination of the gap Δ_i. To be more precise, we are looking for a subset $\mathcal{B}^\star \subset \mathcal{B}_i$, such that by discarding the cutting planes Λ_j $(j \notin \mathcal{B}^\star)$, the gap does not widen. That is, we must have

$$\min_{j\in\mathcal{B}^\star} \left(\Lambda_j(\boldsymbol{x}_j) + \delta_j \right) = U_i \qquad \text{and} \qquad \min_{\boldsymbol{x}\in Q} \left(\max_{j\in\mathcal{B}^\star} \Lambda_j(\boldsymbol{x}) \right) = L_i \,. \tag{8.11}$$

The left-hand equality can be achived by including a single subscript value in $\mathcal{B}^\star$. The right-hand equality can be achived by including at most $K+1$ subscript values. Thus the cardinality of the reduced bundle need not be larger then $K+2$.

We reconstruct the reduced bundle. I.e., for each $j \in \mathcal{B}^\star$, we construct the cutting plane $\Lambda_j^\star$ at the point $\boldsymbol{x}_j$ with accuracy $\delta^\star$ and reliability $p^\star$. (The prescribed accuracy, $\delta^\star$ is computed from the latest safe upper estimate on the gap.) Then we compute

$$U^\star = \min_{j\in\mathcal{B}^\star} \left(\Lambda_j^\star(\boldsymbol{x}_j) + \delta^\star \right), \qquad L^\star = \min_{\boldsymbol{x}\in Q} \left(\max_{j\in\mathcal{B}^\star} \Lambda_j^\star(\boldsymbol{x}) \right), \qquad \Delta^\star = U^\star - L^\star \,. \tag{8.12}$$

The recomputed gap $\Delta^\star$ may be larger then the original Δ_i. However,

$$\Delta^\star < \left(1 - \frac{1}{2}\lambda\right) \Delta_s^{\#} \tag{8.13}$$

holds with a positive probability. (Proposition 36.)

If (8.13) holds then the current bundle is accepted as safe. If (8.13) fails, then we recourse to the latest safe bundle.

Starting from the latest accepted safe bundle, the procedure accepts a new safe bundle with probability 1. (Corollary 37.)

The Framework of the Procedure

Initialization.

Set the stopping tolerance $\epsilon > 0$.

Set the required levels of reliability p, $p^\star$ $\quad (0 < p < p^\star < 1)$.

Set the parameters λ, γ $\quad \left(0 < \lambda < 1, \;\; 0 < \gamma < \min\left\{ (1-\lambda)^2, \, \frac{1}{16}\lambda \right\} \right)$.

Let $\mathcal{B}_0 := \emptyset$. (Start with empty bundle.)

Find a starting point $\boldsymbol{x}_1 \in Q$.

Set starting accuracy $\delta_1 > 0$.

Let $i := 1$. (Iteration counter.)

Set starting *safe upper estimate on the gap* $\Delta_0^\# > 0$.
(A good setting is for example, $\Delta_0^\# := \kappa D + \delta_1$.)
Let $\mathcal{B}_0^\# := \emptyset$. (Starting *accepted safe bundle* is the empty set.)
Let $s := 0$. (Counter of accepted safe bundles.)
Let $i_s^\# := 0$. (Iteration counter value that belongs to sth accepted safe bundle.)

Construct new cutting plane.

Construct the cutting plane Λ_i at the point $\boldsymbol{x}_i$ with accuracy δ_i and reliability p.
Add the new cutting plane to the bundle: $\mathcal{B}_i := \mathcal{B}_{i-1} \cup \{i\}$.

If the consistency check (8.8) fails, then $\rightarrow$*Recourse to latest safe bundle.*

Define the upper cover f_i as in (8.9), and compute U_i, L_i, and Δ_i as in (8.10).
If $\Delta_i < (1-\lambda)\Delta_s^\#$, then $\rightarrow$*Reduce bundle.*

Find new iterate.

Let $\boldsymbol{x}_{i+1}$ be the projection of $\boldsymbol{x}_i$ to the level set $\{\, \boldsymbol{x} \in Q \mid f_i(\boldsymbol{x}) \le L_i + \lambda\Delta_i \,\}$.
Let $\delta_{i+1} := \gamma\Delta_s^\#$.
Increment i.
$\rightarrow$*Construct new cutting plane.*

Reduce bundle.

Determine $\mathcal{B}^\star \subset \mathcal{B}_i$ such that (8.11) holds, and $|\mathcal{B}^\star| \le K+2$.
Let $\delta^\star := \gamma\left(1 - \frac{1}{2}\lambda\right)\Delta_s^\#$.
For each $j \in \mathcal{B}^\star$,
construct the cutting plane $\Lambda_j^\star$ at the point $\boldsymbol{x}_j$ with accuracy $\delta^\star$ and reliability $p^\star$.
(If the reduced bundle should be inconsistent, then this step is repeated.)

Compute $U^\star$, $L^\star$, and $\Delta^\star$ as in (8.12).
If $\Delta^\star \le \left(1 - \frac{1}{2}\lambda\right)\Delta_s^\#$ does not hold (that is, test (8.13) fails),
then $\rightarrow$*Recourse to latest safe bundle.*

For each $j \in \mathcal{B}^\star$, let $\delta_j := \delta^\star$ and $\Lambda_j := \Lambda_j^\star$.

Let $\mathcal{B}_{s+1}^\# := \mathcal{B}^\star$. (Accept reduced bundle as safe.)
Let $\Delta_{s+1}^\# := \left(1 - \frac{1}{2}\lambda\right) \Delta_s^\#$. (Note that we have $\delta_j = \gamma \Delta_{s+1}^\#$ $(j \in \mathcal{B}_{s+1}^\#)$.)
Let $i_{s+1}^\# := i$.
Increment s.

Let $\mathcal{B}_i := \mathcal{B}^\star$. (Reset current bundle.)
Let $U_i := U^\star$, $L_i := L^\star$, and $\Delta_i := \Delta^\star$.

If $\Delta_s^\# \leq \epsilon$ then stop.
$\rightarrow$*Find new iterate.*

Recourse to latest safe bundle.
Let $i := i_s^\#$.
Let $\mathcal{B}_i := \mathcal{B}_s^\#$.

Define the upper cover f_i as in (8.9), and compute U_i, L_i, and Δ_i as in (8.10).

$\rightarrow$*Find new iterate.*

Formal Proofs

Proposition 35 *Suppose that the procedure has just accepted the sth safe bundle, or has just recoursed to the sth safe bundle. Suppose that the stopping condition is not satisfied (I.e., $\Delta_s^\# > \epsilon$).*

If no inconsistency of the kind of (8.8) *is detected in the course of the coming iterations, then the procedure finds a critical iteration within*

$$I \;=\; 1 \;+\; \left(\frac{\kappa D}{[\,(1-\lambda)^2 - \gamma\,]\;\epsilon} \right)^2 \tag{8.14}$$

iterations.

(Inconsistency of the kind of (8.8) is checked by the test performed in paragraph '*Construct new cutting plane*' of the framework.)

Proof. Suppose that we have just performed the paragraph '*Construct new cutting plane*' of the framework, and the iteration counter has the value $i > i_s^{\#}$. Since the consistency check (8.8) was passed, we must have $\Delta_i > 0$.

Hence Proposition 32 is applicable to the bundle $\{ \Lambda_j \ \ (j \in \mathcal{B}_i) \}$. (The proof of this proposition does not require the cutting planes $\Lambda_j \ \ (j \in \mathcal{B}_i)$ to be supporting hyperplanes of a convex function. The only requirement is that the gap must always be non-negative.)

The present statement is just a rewording of Theorem 34. □

Assume that in the sth accepted safe bundle, each cutting plane is well-generated. Moreover, assume that in a sequence of I iterations starting from this safe bundle, each cutting plane is well-generated. In this case, a critical iteration is bound to occur.

Remark. According to the reports in Lemaréchal, Nemirovskii, and Nesterov (1995), we may set $I = K$.

Proposition 36 *Suppose that the procedure has just accepted the sth safe bundle, or has just recoursed to the sth safe bundle. Suppose that the stopping condition is not satisfied.*

Assume that in the sth accepted safe bundle, each cutting plane is well-generated.

Then the event that a new safe bundle is accepted before 'Recourse to the latest safe bundle' becomes necessary, has a probability at least $(p^\star)^{K+2}\, p^I$.

Proof. Starting from the sth safe bundle, the procedure either finds a critical iteration, or recourses to the latest safe bundle in at most I iterations. In case a critical iteration was found, the bundle is reduced and reconstructed. Finally, the test (8.13):

$$\Delta^\star \ < \ \left(1 - \frac{1}{2}\lambda\right) \Delta_s^{\#}$$

is performed (paragraph '*Reduce bundle*' of the framework). A new safe bundle is accepted if the above test is passed. Hence an attempt consists of

– Performing at most I iterations.

– Determining $\mathcal{B}^\star$, and constructing $\left\{ \Lambda_j^\star \ \ (j \in \mathcal{B}^\star) \right\}$, in case a critical iteration was found.

Assume that in the course of an attempt, each cutting plane is well-generated. We prove that the test (8.13) is bound to be passed in this case.

A critical iteration is bound to occur under the assumption. Suppose that the $(i-1)$th iteration is found critical.

From the assumption, it also follows that

$$|\Lambda_j(\boldsymbol{x}) - \Lambda_j^\star(\boldsymbol{x})| \ \leq \ 2(\delta_j + \delta^\star) \quad (\boldsymbol{x} \in Q) \qquad (j \in \mathcal{B}^\star)\,. \tag{8.15}$$

Thus we have

$$|\Lambda_j(\boldsymbol{x}) - \Lambda_j^\star(\boldsymbol{x})| \ \leq \ 2(\delta_j + \delta^\star) \ = \ 2\left(\gamma\Delta_s^{\#} + \gamma\left(1 - \tfrac{1}{2}\lambda\right)\Delta_s^{\#}\right) \ = \ 4\gamma\left(1 - \tfrac{1}{4}\lambda\right)\Delta_s^{\#}$$

$$\leq \ 4\gamma\left(1 - \tfrac{1}{4}\lambda\right)\Delta_s^{\#} \qquad (\boldsymbol{x} \in Q, \quad j \in \mathcal{B}^\star)\,.$$

It follows that

$$U^\star - U_i \;\leq\; 4\gamma\left(1-\frac{1}{4}\lambda\right)\Delta_s^\# \qquad \text{and} \qquad L_i - L^\star \;\leq\; 4\gamma\left(1-\frac{1}{4}\lambda\right)\Delta_s^\#\,;$$

and hence $\Delta^\star - \Delta_i \;\leq\; 8\gamma\left(1-\frac{1}{4}\lambda\right)\Delta_s^\#$. We also have $\Delta_i \;\leq\; (1-\lambda)\Delta_s^\#$, because the $(i-1)$th iteration is critical. Adding up these two inequalities, we get

$$\Delta^\star \;\leq\; \left[\,8\gamma\left(1-\frac{1}{4}\lambda\right) \;+\; (1-\lambda)\,\right]\,\Delta_s^\# \;\leq\; [\,8\gamma \;+\; (1-\lambda)\,]\,\Delta_s^\#\,.$$

Using $\gamma \leq \frac{1}{16}\lambda$, it follows that the test (8.13) is passed.

Hence the probability of passing the test is larger or equal to the probability of the event that *each cutting plane is well-generated in the course of the attempt.* The latter probability is at least

$$p^I\;(p^\star)^{K+2}\,.$$

□

Corollary 37 *Suppose that the procedure has just accepted the sth safe bundle, or has just recoursed to the sth safe bundle. Suppose that the stopping condition is not satisfied.*

Assume that sth safe bundle is well-generated. (I.e., each cutting plane is well-generated in it.)

Then the procedure will accept a new, $(s+1)$th safe bundle with probability 1.

Now let us compute the conditional probability that the $(s+1)$th safe bundle is well-generated, given that the sth safe bundle is well-generated. It will be somewhat less then $(p^\star)^{K+2}$, because there is a decision involved.

Starting from the sth safe bundle, an attempt to find a new safe bundle consists of

- Performing at most I iterations. The probability that each cutting planes is well-generated is at least $q = p^I$.
- Determining $\mathcal{B}^\star$, and constructing $\left\{\; \Lambda_j^\star \;\; (j \in \mathcal{B}^\star)\;\right\}$, in case a critical iteration was found. The probability that each cutting plane is well-generated is at least $q^\star = (p^\star)^{K+2}$.

The outcome of a single attempt can be

- accepting a well-generated bundle. The probability of this event is at least $r^+ = qq^\star$.
- or accepting a not well-generated bundle. The probability of this event is at most $r^- = (1-q^\star)$.
- or recoursing to the sth safe bundle.

As the attempt is iterated, a new bundle will be accepted with probability 1. The probability that a well-generated bundle is accepted is at least

$$r = r^+ + (1 - r^+ - r^-)r^+ + (1 - r^+ - r^-)^2 r^+ + \ldots = \frac{qq^\star}{qq^\star + (1 - q^\star)}.$$

According to Theorem 34, the number of safe bundles accepted in the course of the procedure need not be larger then $I^\# = \tilde{C} \log\left(\frac{\kappa D}{\epsilon}\right) + 1$ with $\tilde{C} = 1 \Big/ -\log(1 - \lambda)$.

Hence the probability of terminating with near-optimal solution is at least

$$r^{I^\#}.$$

Improvements on the Procedure

A drawback of the procedure is that if erroneous data are accepted in a safe bundle, then these data may never be corrected. If the procedure can not find a near-optimal solution within $I^\#$ critical iterations, then we better stop it.

On the other hand, the probability of finding a near-optimal solution can be increased by iterating the procedure.

8.3 Application of the Stochastic Level Method to the Dynamic Problem

We wish to minimize $f = \Theta$ over $Q = P$. In course of the method, we need to minimize linear and convex separable quadratic functions over the level sets

$$Q_i = \{ \boldsymbol{x} \in Q \mid f_i(\boldsymbol{x}) \leq L_i + \lambda\Delta_i \}.$$

The constraints $f_i(\boldsymbol{x}) \leq L_i + \lambda\Delta_i$ are known linear functions, thus easy to handle. The polyhedron $Q = P$, on the other hand, is not known explicitly, but can be explored by the method of Gilmore and Gomory.

The linear and quadratic programming subproblems can be solved by the procedures proposed in Chapters 3 and 4.

Magnitudes in a Realistic Case

We have $H = 75$ and $m = 3$; $l_1 = 3$, $l_2 = 5$, $l_3 = 7$. Hence the dimension is $K = m\lfloor H \rfloor = 225$. Moreover, we have

$$D = \mathrm{Diam}(P) \leq 34.$$

Assume that $\boldsymbol{q}' = \boldsymbol{q}$, i.e., the prices of different products are the same in each production period.

Let us consider the highest profit that the manufacturer can make on a unit length of fiber sold. That is

$$\bar{q} = \max_{1 \le k \le m} \frac{q_k}{l_k} .$$

Let moreover $\tilde{q}$ denote the average profit the manufacturer usually makes on a unit length of fiber produced. Assume that $\bar{q} \le 2\tilde{q}$.

The accuracy required from the final solution is one percent of the expected profit to be realized in a single production period. That is

$$\epsilon \;=\; 0.01\, \tilde{q}\, HN .$$

(N denotes the number of preforms processed during a single period.)

For the estimation of the Lipschitz constant κ, let us use the expression derived in Chapter 7, in paragraph **Lipschitz Condition** (just before Section 7.1).

$$\kappa^2 \;\approx\; \sum_{k,\,h} \;\left(q_k \,\mathrm{E}(|\phi_h{}'|) \right)^2 \;\approx\; \sum_{k=1}^{m} \; q_k^2 \sum_{h=1}^{\lfloor H \rfloor} \left(\mathrm{E}(|\phi_h{}'|) \right)^2 \;\le\; \sum_{k=1}^{m} \; q_k^2 \left(\sum_{h=1}^{\lfloor H \rfloor} \mathrm{E}(|\phi_h{}'|) \right)^2$$

$$\approx\; \sum_{k=1}^{m} \; q_k^2 \,(\,NH)^2 \;=\; (\,NH)^2 \sum_{k=1}^{m} \; q_k^2 .$$

Hence we get

$$\kappa \;\le\; NH \sqrt{\sum_{k=1}^{m} \; q_k^2} \;\le\; NH \sqrt{m} \max_k q_k \;\le\; NH \sqrt{m}\, \bar{q} H .$$

Thus

$$\frac{\kappa D}{\epsilon} \;\le\; \frac{NH \sqrt{m}\, \bar{q} H\, D}{0.01\, \tilde{q} HN} \;=\; 100\, HD \sqrt{m}\, \frac{\bar{q}}{\tilde{q}} \;\le\; 10^6 .$$

And so

$$\log \left(\frac{\kappa D}{\epsilon} \right) \;\le\; 20 .$$

Setting the parameter λ to 0.5, the constant $\tilde{C}$ is 1. Hence the number of critical iterations is not larger then $I^{\#} \;=\; 21$.

Let us set the reliability parameters to $p = p^\star = 0.99994$ (consider the Remarks at the end of Section 7.2.1). Assume that the number of critical iterations between two subsequent critical iterations does not exceed $I = K$. Then we have

$$q = p^K \approx 0.987, \quad q^\star = (p^\star)^{K+2} \approx 0.987,$$

$$r = \tfrac{qq^\star}{qq^\star + (1-q^\star)} \approx 0.987,$$

$$r^{I^{\#}} \approx 0.77 .$$

Bibliography

[1] Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis.* John Wiley, New York.

[2] Dantzig, G.B., and Ph. Wolfe (1961). The Decomposition Algorithm for Linear Programs. *Econometrica* **29**, 767-778.

[3] Deák, I. (1990). *Random Number Generators and Simulation.* Akadémiai Kiadó, Budapest.

[4] Fábián, Cs.I. (1999). Az üvegszál anyaghibáinak valószínűségi modellezése. *Alkalmazott Matematikai Lapok* **19**, 91-100.

[5] Fábián, Cs.I. (1998). Stochastic Programming Model for Optical Fiber Manufacturing. *RUTCOR Research Report* **34-98**.

[6] Fábián, Cs.I., A. Prékopa, and O. Ruf-Fiedler (1999). On a Dual Method for a Specially Structured Linear Programming Problem. *RUTCOR Research Report* **3-99**.

[7] Gilmore, P.C. and R.E. Gomory (1961). A Linear Programming Approach to the Cutting Stock Problem. *Operations Research* **9**, 849-858.

[8] Hiriart-Urruty, J-B. and C. Lemaréchal. (1993). *Convex Analysis and Minimization Algorithms.* Volume II: *Advanced Theory and Bundle Methods.* Springer, Berlin.

[9] Kelley, J.E. (1960). The Cutting Plane Method for Solving Convex Programs. *Journal of the SIAM* **8**, 703-712.

[10] Lemaréchal, C., A. Nemirovskii, and Yu. Nesterov. (1995). New Variants of Bundle Methods. *Mathematical Programming* **69**, 111-147.

[11] Markowitz, H. (1952). Portfolio Selection. *Journal of Finance* **7**, 77-91.

[12] Murr, M.R. and A. Prékopa (1996). Solution of a Product Substitution Problem Using Stochastic Programming. *RUTCOR Research Report* **32-96**.

[13] PRÉKOPA, A. (1990). Dual Method for the Solution of a One-Stage Stochastic Programming Problem with Random RHS Obeying a Discrete Probability Distribution. *ZOR* **34**, 441-461.

[14] PRÉKOPA, A. (1995). *Stochastic Programming.* Akadémiai Kiadó, Budapest and Kluwer Academic Publishers, Dordrecht.

[15] ROCKAFELLAR, R.T. (1970). *Convex Analysis.* Princeton University Press, Princeton.

[16] SZÁNTAI, T. (1988). A Computer Code for the Solution of Probabilistic-constrained Stochastic Programming Problems. In *Numerical Techniques for Stochastic Optimization*, Yu. Ermoliev and R.J-B. Wets (eds.). Springer-Verlag, New York.

[17] ŠIDAK, Z. (1968). On Multivariate Normal Probabilities of Rectangles: Their Dependence on Correlations. *The Annals of Mathematical Statistics* **39**, 1425-1434.

[18] WILKS, S.S. (1962). *Mathematical Statistics.* John Wiley, New York.

[19] ZOUTENDIJK, G. (1960). *Methods of Feasible Directions.* Elsevier, Amsterdam.