

## Chapter 2

# Single-stage SLP models

### 2.1 Introduction

In this chapter we consider stochastic programming problems which represent a single decision stage. The decision is to be made “here and now” and the models do not account for any corrective (recourse) actions which might be available after the realization of the random variables in the model becomes known. Such type of models typically involve, either in the constraints or in the objective function, or in both of them, random variables of the following form

$$\zeta(x, \xi) := T(\xi)x - h(\xi) \quad (2.1)$$

where  $\xi : \Omega \rightarrow \mathbb{R}^r$  is a random vector on a probability space  $(\Omega, \mathcal{F}, P)$ .  $T(\xi)$  denotes a random  $s \times n$  matrix,  $h(\xi) \in \mathbb{R}^s$  stands for a random vector, both depending on the random vector  $\xi$ . The support of  $\xi$  is defined as the smallest closed set  $\Xi \subset \mathbb{R}^r$  having the property  $\mathbb{P}(\xi \in \Xi) = 1$ .

For being more specific, we assume that the dependence is defined in terms of affine linear relations as follows: for all  $\xi \in \Xi$  we have

$$\begin{aligned} T(\xi) &= T + \sum_{j=1}^r T_j \xi_j, \\ h(\xi) &= h + \sum_{j=1}^r h_j \xi_j, \end{aligned} \quad (2.2)$$

where  $T, T_j \in \mathbb{R}^{s \times n}$  are deterministic matrices and  $h, h_j \in \mathbb{R}^s$  are deterministic vectors,  $j = 1, \dots, r$ .

In this chapter the particular form (2.2) will not be used explicitly. All we need is the joint probability distribution of  $(T(\xi), h(\xi))$  which will be presupposed as known throughout. As for stochastic programming in general, the basic assumption is that the probability distribution of  $(T(\xi), h(\xi))$  does not depend on  $x$ . This means

that our decision has no influence on the probability distribution of the random entries in the model data.

If in the constraints,  $\zeta(x, \xi)$  frequently plays the role of a random slack variable in a random linear inequality. For instance, taking the inequality  $T(\xi)x \geq h(\xi)$ , this inequality can evidently also be written in the form  $\zeta(x, \xi) \geq 0$ .

For later reference, we write (2.1) also in a row-wise form as

$$\zeta_i(x; \xi) := t_i^T(\xi)x - h_i(\xi), \quad i = 1, \dots, s, \quad (2.3)$$

where the components of the  $n$ -dimensional random vector  $t_i(\xi)$  are the elements of the  $i^{\text{th}}$  row of  $T(\xi)$ ,  $i = 1, \dots, s$ . Alternatively, (2.1), may be written in a column-wise fashion as

$$\zeta(x, \xi) = \sum_{j=1}^n T_j(\xi)x_j - h(\xi), \quad (2.4)$$

where the  $s$ -dimensional random vector  $T_j(\xi)$  denotes the  $j^{\text{th}}$  column of  $T(\xi)$ ,  $j = 1, \dots, n$ . Thus  $\zeta(x, \xi)$  can be regarded as an affine linear combination of random vectors. Our assumption is that the joint probability distribution of these random vectors is known. The coefficients in the linear combination are the decision variables  $x_j$ , consequently the probability distribution of  $\zeta(x, \xi)$  will depend on our decision. We control the probability distribution of  $\zeta(x, \xi)$ , by controlling its realizations, according to (2.4).

The question arises, what can be stated about the probability distribution of  $\zeta(x, \xi)$ ? In particular, assuming that the joint probability distribution of  $(T_j(\xi), j = 1, \dots, n; h(\xi))$  belongs to a given parametric family of distributions, for which families will the affine linear combination  $\zeta(x, \xi)$  belong to the same family? An example of a family, for which the answer is affirmative, is the class of multivariate normal distributions. This question will be further pursued in Section 2.2.3, in connection with separate probability constraints.

Note that a similar question also arises in mathematical statistics regarding linear statistical models. In that case  $h(\xi)$  represents an error (noise) term, which is usually assumed as being stochastically independent of the random vectors  $T_j(\xi)$ . In mathematical statistics we are dealing with a random vector  $\zeta$  with unknown distribution and the goal is to choose  $x$  in such a way, that the distribution of  $\zeta(x, \xi)$  provides a good approximation to the distribution of  $\zeta$  in a statistical sense. For achieving this, the  $x_j$ 's are considered as random variables. The starting point is a joint sample according to the distribution of  $(\zeta, T_j(\xi); j = 1, \dots, n)$  and assuming the linear model (2.4), the aim is to construct unbiased estimators for the  $x_j$ 's.

In stochastic programming we face a different situation. The primary entity is the given joint distribution of  $(T_j(\xi), j = 1, \dots, n; h(\xi))$  and the goal is to achieve a probability distribution of  $\zeta(x, \xi)$  with advantageous properties, whereby  $x$  is considered as being deterministic. To make this precise, we will attach a quantitative meaning to the term "advantageous" and will arrive this way at a classification scheme for the different classes of SLP models as follows:

- First we define a function  $\rho : \Upsilon \rightarrow \mathbb{R}^1$  for evaluating random vectors, where  $\Upsilon$  is some linear space of  $s$ -dimensional random vectors defined on a probability space  $(\Omega, \mathcal{F}, P)$ . For instance,  $\Upsilon$  will be frequently chosen as the linear space of random vectors with finite expected value. For each random vector  $\vartheta \in \Upsilon$ ,  $\rho(\vartheta)$  is interpreted as a quality measure in the corresponding modeling approach. Depending on the interpretation of  $\rho(\vartheta)$  as either expressing opportunity or risk, “advantageous” will mean that higher or lower values of  $\rho(\vartheta)$  are considered as preferable, respectively. In the latter case  $\rho$  will be called a risk measure. The probability distribution function of  $\vartheta$  will be denoted by  $F_\vartheta$  and  $\Theta$  will denote the support of  $\vartheta$ . In the special case  $s = 1$ ,  $\Upsilon$  is some linear space of random variables. In the sequel, the term random vector will always mean that  $s > 1$  is permitted whereas the term random variable will indicate that  $s = 1$  is assumed.
- Based on the chosen function  $\rho$  for evaluating random variables, decision vectors  $x$  will be evaluated as follows. We define the corresponding evaluation function  $V : \mathbb{R}^n \rightarrow \mathbb{R}^1$  by substituting  $\zeta(x, \xi)$  into  $\rho$ :

$$V(x) := \rho(\zeta(x, \xi)) \quad (2.5)$$

provided that  $\zeta(x, \xi) \in \Upsilon$  holds for all  $x$ .  $V(x)$  will be interpreted as a quality measure for  $x$  and will be employed for building SLP models. For indicating that the evaluation involves all components of the random vector simultaneously, we will call  $V$  a *joint evaluation function*.

- Alternatively, when dealing with constraints, it may make sense to assign quality measures to the components of  $\zeta(x, \xi)$  separately. If  $\rho$  is defined for random variables and  $\zeta_i(x, \xi) \in \Upsilon$  holds for all  $x$  and all  $i$  then  $V_i(x) := \rho(\zeta_i(x, \xi))$  serves for evaluating  $x$  for the  $i^{\text{th}}$  component of  $\zeta(x, \xi)$ ,  $i = 1, \dots, s$ . Concerning  $V_i$ , the term *separate evaluation function* will be employed, for pointing out the fact that the components of the random vector  $\zeta(x, \xi)$  are evaluated separately. If  $s = 1$  holds then  $\zeta(x, \xi)$  is a random variable and both adjectives “separate” and “joint” apply. This ambiguity will have no substantial influence on the discussions concerning SLP models.

Having chosen  $\rho$ , the evaluation function  $V$  is uniquely defined. The different SLP model classes will correspond to different choices of the quality measure  $\rho$  for random vectors.

$\Upsilon$  will be one of the following linear spaces of  $s$ -dimensional random vectors:

$$\begin{aligned} \mathcal{L}_s^0 &:= \{ \text{the set of all random vectors on } (\Omega, \mathcal{F}, P) \}, \\ \mathcal{L}_s^1 &:= \mathcal{L}_s^1(\Omega, \mathcal{F}, P) = \{ \vartheta \mid \int_{\mathbb{R}^s} \|t\|_1 dF_\vartheta(t) < +\infty \}, \\ \mathcal{L}_s^2 &:= \mathcal{L}_s^2(\Omega, \mathcal{F}, P) = \{ \vartheta \mid \int_{\mathbb{R}^s} \|t\|_2^2 dF_\vartheta(t) < +\infty \}, \\ \mathcal{L}_s^\infty &:= \mathcal{L}_s^\infty(\Omega, \mathcal{F}, P) = \{ \vartheta \mid \exists C : \mathbb{P}(\|\vartheta\|_2 > C) = 0 \}, \end{aligned} \quad (2.6)$$

where  $\mathcal{L}_s^1$  is the space of  $s$ -dimensional random vectors with finite expected value,  $\mathcal{L}_s^2$  stands for the space of random vectors with finite second moments, and  $\mathcal{L}_s^\infty$  denotes the space of random vectors having a bounded support.  $\|t\|_2 = \sqrt{\sum_{i=1}^s t_i^2}$  is the Euclidean norm and  $\|t\|_1 = \sum_{i=1}^s |t_i|$  holds.

Note that up to this point we have viewed  $\zeta(x, \xi) = T(\xi)x - h(\xi)$  as an affine linear combination of random vectors. Alternatively, we can also consider  $\zeta(x, \xi)$  as a *deviation* between  $T(\xi)x$  and  $h(\xi)$ . In mathematical statistics an interpretation could be fitting  $T(\xi)x$  to  $h(\xi)$  in a least squares sense. In this setting,  $\zeta(x, \xi)$  would be an error term. Assuming some distributional properties of the error term and having a sample for  $(T(\xi), h(\xi))$ , the goal in mathematical statistics is to find a good fit. In stochastic programming we proceed analogously as before: quality measures for random variables will be introduced and stochastic programming models will be built by employing the corresponding evaluation function  $V$ . We interpret the quality measure in this case as *deviation measure*.

As mentioned above, SLP models will be built by employing evaluation functions  $V$  corresponding to some quality measure  $\rho$ . The different SLP model classes will be discussed in a framework of prototype models. For employing joint- and separate evaluation functions in the constraints, we consider the models

$$\left. \begin{array}{l} \max c^T x \\ \text{s.t. } V(x) \geq \kappa \\ x \in \mathcal{B} \end{array} \right\} \quad \left. \begin{array}{l} \max c^T x \\ \text{s.t. } V_i(x) \geq \kappa_i, \\ \quad i = 1, \dots, s \\ x \in \mathcal{B}, \end{array} \right\} \quad (2.7)$$

where  $\kappa$  and  $\kappa_i$  are prescribed,  $i = 1, \dots, s$ , and  $\mathcal{B}$  is a polyhedral set

$$\mathcal{B} = \{x \mid Ax \preceq b, l \leq x \leq u\} \quad (2.8)$$

with  $A$  being an  $m \times n$  matrix and  $x$ ,  $b$ ,  $l$ , and  $u$  having corresponding dimensions. The symbol  $\preceq$  means that any one of the relations  $\leq$ ,  $=$ , and  $\geq$  is permitted row-wise.

For models with the evaluation function being in the objective, we consider the prototype model

$$\left. \begin{array}{l} \max c^T x + V(x) \\ \text{s.t. } x \in \mathcal{B}. \end{array} \right\} \quad (2.9)$$

Alternatively, we will also employ prototype models with reversed direction of the inequalities in the constraints of (2.7) and with minimization instead of maximization in (2.9). To see the reason for this, let us assume first, that for some model class the evaluation function  $V$  is a concave function. In this case, both (2.7) and (2.9) are convex programming problems. Assume next that for some other SLP model class  $V$  turns out to be a nonlinear convex function. In this case our prototype models become non-convex optimization problems, whereas their counterparts with reversed inequality constraints and minimization in the objective will be convex programming problems. The point is that the chances for finding efficient algorithms are

much better for convex optimization problems than for the non-convex case. This subject will be further pursued in Section 2.6.

From the modeling viewpoint, stochastic programming models can have a composite form, involving several different random vectors of the type (2.1). We have chosen to work with the above prototype models because they serve well for explaining the basic ideas which can then be applied to composite models in a straightforward way. For some model classes  $c = 0$  will be required in (2.9). The reason is that, for those model classes,  $V$  has merely some generalized concavity property which might be destroyed by adding a linear term.

The objective function of (2.9) consists of a sum of two terms whereas in applications they are usually weighted with respect to each other. Weighting can also be interpreted in terms of duality. We take as an example the following weighted version of (2.9):

$$\begin{aligned} \max \quad & c^T x + \lambda V(x) \\ \text{s.t.} \quad & x \in \mathcal{B} \end{aligned} \quad (2.10)$$

with a positive weight  $\lambda$ . This can equivalently be written in the form

$$\begin{aligned} v(\lambda) := \max \quad & c^T x + \lambda (V(x) - \kappa) \\ \text{s.t.} \quad & x \in \mathcal{B} \end{aligned} \quad (2.11)$$

where  $-\lambda \kappa$  is merely a shift in the optimal objective value. This problem is called a *Lagrangian relaxation* of the first optimization problem in (2.7). The corresponding Lagrange-dual-problem is then

$$\min \{ v(\lambda) \mid \lambda \geq 0 \}. \quad (2.12)$$

For the duality relationships between (2.7), (2.11), and (2.12) see Bazaraa and Shetty [11].

For the sake of simplicity of presentation, we assume in the sequel that positive weighting factors (if any) are taken into account in the definition of  $c$ .

The simplest way for assigning a quality measure to  $\zeta(x, \xi)$  is taking the expectation. To see how this works, let us discuss the application of the idea for including a system of random inequalities  $\zeta(x, \xi) \geq 0$  into an SLP model. We choose separate evaluation for the components of  $\zeta(x, \xi)$  and employ the quality measure

$$\rho_E(\vartheta) := \mathbb{E}[\vartheta], \quad \vartheta \in \mathcal{L}_1^1 \quad (2.13)$$

for the components. Assuming the existence of the expected values of  $T(\xi)$  and  $h(\xi)$  and setting  $\kappa_i = 0$  for all  $i$ , this leads to the following formulation of (2.7):

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & \bar{t}_i^T x \geq \bar{h}_i, \quad i = 1, \dots, n \\ & x \in \mathcal{B} \end{aligned} \quad (2.14)$$

where  $\bar{t}_i := \mathbb{E}_\xi[t_i(\xi)]$  and  $\bar{h}_i := \mathbb{E}_\xi[h_i(\xi)]$  hold, with the components of  $t_i(\xi)$  being the elements of the  $i^{\text{th}}$  row of  $T(\xi)$ . The resulting deterministic LP problem is called *expected value problem*. Unfortunately, the expected value problem is frequently used as a substitute for the SLP problem. While in some (rare) situations this might be appropriate, in general it is a very crude approach: the whole probability distribution is collapsed into a one-point distribution. It should by no means be used as the single way for representing  $\zeta(x, \xi)$  in the model. However, accompanied by a constraint or objective part involving some other quality measure, it can prove to be an important constituent of the SLP model. For examples of this kind see Section 2.7.3. In financial portfolio optimization, the most prominent and widely used model of the combined type is the model of Markowitz [217], see also Elton et al. [85].

For discussing the next idea, our starting point is again the system of random inequalities  $\zeta(x, \xi) \geq 0$ . We interpret this as prescribing the sign of  $\zeta(x, \xi)$  and consider the inclusion of the system of random inequalities

$$T(\xi)x \geq h(\xi) \quad (2.15)$$

into the stochastic programming model. The difficulty is that, besides the decision vector  $x$ , the constraints also depend on the random vector  $\xi$ . One of the earliest proposals for overcoming this difficulty is due to Madansky [212], [213], who suggested a worst-case approach by prescribing the inequalities (2.15) for all  $\xi \in \Xi$ , with  $\Xi$  denoting the support of the random vector  $\xi$ . We assume that  $\Xi$  is a bounded set. This leads to the following formulation of (2.7):

$$\left. \begin{array}{ll} \max & c^T x \\ \text{s.t.} & T(\xi)x \geq h(\xi), \quad \xi \in \Xi \\ & x \in \mathcal{B}. \end{array} \right\} \quad (2.16)$$

Madansky termed the solution of this optimization problem as *fat solution*. The approach corresponds to the following choice of the quality measure  $\rho_{\text{fat}}: \Upsilon = \mathcal{L}_s^\infty$  is the set of random vectors having a bounded support and

$$\rho_{\text{fat}}(\vartheta) := \min_{\hat{\vartheta} \in \Theta} \min_{1 \leq i \leq s} \hat{\vartheta}_i, \quad \vartheta \in \mathcal{L}_s^\infty, \quad (2.17)$$

where  $\Theta$  is the support of  $\vartheta$ . The formulation (2.16) corresponds to the model (2.7) with the inequality constraint chosen as  $V(x) = \rho_{\text{fat}}(\zeta(x, \xi)) \geq 0$ . The feasible domain  $\mathcal{D}$  of (2.16) is the intersection of convex sets and thus it is obviously convex:

$$\mathcal{D} = \bigcap_{\xi \in \Xi} \{x \mid T(\xi)x \geq h(\xi), x \in \mathcal{B}\}.$$

In the special case of a finite discrete distribution,  $\Xi$  is a finite set and (2.16) reduces to a linear programming problem. In general, (2.16) may turn out in many cases as being infeasible, especially if  $\Xi$  contains infinitely many points.

Recently, after the new optimization area of semidefinite programming has emerged in the 1990s, it became numerically feasible to compute fat solutions also for bounded domains  $\Xi$  containing infinitely many points. The idea is that instead of considering  $\Xi$  as an index set,  $\xi \in \Xi$  is explicitly handled as a constraint in (2.16) and  $\xi$  is considered as a deterministic variable. For instance, with ellipsoidal domains  $\Xi$ , (2.16) can be reformulated as an equivalent semidefinite programming problem, see Ben–Tal et al. [15] and the references therein. The cited paper also presents an extension of this approach to the class of semidefinite programming problems. Along with the extension, the approach has also been renamed as *robust optimization*. There are important application areas where working with fat solutions makes sense. As an example, let us mention structural design for mechanical structures, see Ben–Tal et al. [15]. Note that the term “robust optimization” is also used for other model classes; we will return to this point later.

Although in robust optimization, as defined above,  $\Xi$  is called the domain of uncertainty, the approach is only loosely connected to stochastic programming or to stochastic modeling in general. It can be considered as a kind of worst–case parametric programming approach. If, as in our case,  $\Xi$  is the support of a random variable  $\xi$ , the probability distribution of  $\xi$  does not play any role: the models will deliver identical results for all random variables having the same support. For these reasons, the topic of the above kind of robust optimization will not be pursued further in this book. Let us emphasize, however, that robust optimization is an important alternative modeling approach for dealing with uncertain data. For the interested reader we recommend the recent book of Ben–Tal et al. [16] and the references therein.

A straightforward idea for generalizing (2.16) is to consider  $x$  as a feasible solution, if it satisfies all random inequalities for restricted subsets of the support. A natural idea for imposing such a restriction is to consider subsets with prescribed probability levels. SLP models of this class have been introduced and first studied by Charnes and Cooper [41], Miller and Wagner [234] and by Prékopa [258]. The corresponding quality measure is

$$\rho_p(\vartheta) := \mathbb{P}(\vartheta \geq 0), \quad \vartheta \in \mathcal{L}_s^0,$$

defined on the set of all random vectors on  $(\Omega, \mathcal{F}, P)$ . The evaluation function  $V(x)$  (see (2.5)) will be denoted for this model class by  $G(x)$ . This leads to the concept of probability functions, defined as follows:

$$G(x) := \mathbb{P}_\xi( T(\xi)x \geq h(\xi) ). \quad (2.18)$$

Choosing constraints of the form  $G(x) \geq \alpha$ , with  $\alpha$  being a high probability level (for instance,  $\alpha = 0.99$ ), the prototype model (2.7) assumes the form

$$\left. \begin{array}{ll} \max & c^T x \\ \text{s.t.} & \mathbb{P}_\xi( T(\xi)x \geq h(\xi) ) \geq \alpha \\ & x \in \mathcal{B}. \end{array} \right\} \quad (2.19)$$

By choosing  $\alpha = 1$  in this model, we obtain a generalization of the concept of a fat solution, discussed on page 76. In this case  $x \in \mathcal{B}$  is considered as feasible, if the random inequalities hold in an almost sure sense, meaning that they hold except for a subset of  $\Omega$  having probability measure zero.

Taking the quality measure separately for the components of  $\zeta(x, \xi)$ , the constraints in (2.7) are  $G_i(x) \geq \alpha_i$ , with the probability functions  $G_i(x) := \mathbb{P}_\xi (t_i^T(\xi)x \geq h_i(\xi))$ . The probability levels  $\alpha_i$  are specified separately for the individual rows.

Being in the objective, the probability function will be maximized.

Alternatively, we might be interested in constraints of the form  $G(x) \leq \beta$ , with  $\beta$  being small (for instance,  $\beta = 0.01$ ). In this context,  $\beta$  frequently represents a ruin probability, meaning, for instance, the probability of financial ruin of a company, death of a patient, or crashing of a bridge. In such modeling situations, (2.9) would be formulated with minimizing  $G$  in the objective.

Constraints involving probability functions are called chance-constraints or probabilistic constraints. Depending on whether  $G(x)$  or  $G_i(x)$ ,  $i = 1, \dots, s$ , is used, the constraints are called *joint* or *separate* constraints, respectively. From another point of view, a separate constraint is a special case of (2.18) with  $T(\xi)$  consisting of a single row ( $s = 1$ ). Models based on probability functions provide a natural way of building models in several application areas, see Prékopa [266]. Here we just point out two fields, where probabilities play an important part in planning anyhow: finance (ruin probability) and electrical power systems engineering (loss-of-load probability (LOLP)). Stochastic optimization problems involving probability functions will be discussed in detail in Section 2.2.

Let us consider a model involving a probability constraint of the form  $G(x) = \mathbb{P}_\xi (T(\xi)x \geq h(\xi)) \geq \alpha$ , with a high probability level  $\alpha$ . For each fixed  $x$  we interpret the event, that some of the random inequalities do not hold, as loss. Such type of models have the following characteristic feature: On the one hand, they ensure that a loss may only occur with a small probability ( $1 - \alpha$ ). On the other hand, losses may occur, and for the case when they occur, the models provide no control for the modeler on the size of the loss. In modeling situations, where considering the size of the loss makes sense at all, the second characteristic might be considered as a drawback. To distinguish between models based on probability constraints and models which account for the loss size, Klein Haneveld [188] calls the quality measure based on probability functions *qualitative* and quality measures accounting also for the loss size *quantitative*.

Let us discuss shortly situations where the size of the loss does not matter. As a hypothetical example let us imagine that a medical treatment is modeled and the random inequalities in (2.18) express the survival of the patient. Loss means in this case that the patient dies and the size of the loss is meaningless in the modeling context. As a more practical example let us consider mechanical truss optimization problems with a given topology. Such models contain several groups of constraints modeling the laws of mechanics. Under random loads these models may involve chance-constraints of the above type (see, for instance, Marti [224] and the references therein). The random inequalities in (2.18) express some mechanical require-



ments; if they do not hold, then the system crashes. The point is that if the system crashes, then the topology obviously changes and the whole model becomes invalid (the model crashes too). Therefore, it is pointless to include constraints accounting for the size of the loss.

For the case when taking into account the loss–size makes sense, several kinds of remedies have been suggested. It is usually assumed that penalty costs are available for the losses. Prékopa [266] proposes a combined model, involving both probabilistic constraints and recourse–constraints in a two stage recourse problem, with the expected penalty costs for the losses included as an additive term into the objective function. Dert [68] introduces besides the probabilistic constraint binary variables for indicating the occurrence of losses and uses a penalty term in the objective function for the expected penalty costs of losses.

For introducing the next model class we assume that negative values of  $\zeta(x, \xi)$  represent losses and positive values correspond to gains. For the sake of simplicity of presentation we also assume that  $\zeta(x, \xi)$  is a random variable ( $s = 1$  holds). The loss as a random variable can then be written as

$$\zeta^-(x, \xi) := (t^T(\xi)x - h(\xi))^-,$$

where  $t(\xi)$  denotes the single row of  $T(\xi)$ ,  $h(\xi)$  is a random variable, and  $z^- = \max\{0, -z\}$  denotes the negative part of  $z \in \mathbb{R}$ .

Using this, the probability constraint  $G(x) \geq \alpha$  can be written in expected–value terms as

$$G(x) \geq \alpha \iff \mathbb{E}_\xi[\chi(\zeta^-(x, \xi))] \leq 1 - \alpha \quad (2.20)$$

with  $\chi$  denoting the indicator function

$$\chi(z) = \begin{cases} 0 & \text{if } z \leq 0, \\ 1 & \text{if } z > 0. \end{cases}$$

In (2.20) the function  $\chi$  enforces equality across different loss–sizes. Due to an idea of Klein Haneveld [188],  $\chi$  is dropped and the following quality measure is introduced:

$$\rho_{\text{sic}}^-(\vartheta) := \mathbb{E}[\vartheta^-], \quad \vartheta \in \mathcal{L}_1^+.$$

This results in an evaluation function  $H(x) := \mathbb{E}_\xi[\zeta^-(x, \xi)]$  which is simply the expected value of the random variable expressing losses. In models based on this evaluation function, constraints of the form  $H(x) \leq \gamma$  will be employed, where  $\gamma$  is a prescribed maximal level of tolerable expected loss. Constraints based on  $H(x)$  are called *integrated chance constraints*. If in the objective,  $H(x)$  will be minimized. The prototype model with integrated chance constraint has the form

$$\left. \begin{array}{l} \min c^T x \\ \text{s.t. } \mathbb{E}_\xi[\zeta^-(x, \xi)] \leq \gamma \\ \quad \quad \quad x \in \mathcal{B}. \end{array} \right\} \quad (2.21)$$

For the integrated chance constraints which we have considered so far, only  $\zeta^-(x, \xi)$  plays a role. It might be desirable to take into account the entire distribution of  $\zeta(x, \xi)$ . In fact, the following variant of integrated chance constraints takes into account also the expected gain  $\zeta^+(x, \xi)$ :

$$\mathbb{E}_\xi[\zeta^-(x, \xi)] \leq \alpha \mathbb{E}_\xi[|\zeta(x, \xi)|]$$

which can be derived from the quality measure

$$\rho_{\text{sicm}}^\alpha(\vartheta) := (1 - \alpha)\mathbb{E}[\vartheta^-] - \alpha\mathbb{E}[\vartheta^+], \quad \vartheta \in \mathcal{L}_1^1 \quad (2.22)$$

and leads to a convex programming formulation for  $\alpha \leq \frac{1}{2}$ . Integrated chance constraints, including joint constraints for the case when  $\zeta(x, \xi)$  is a random vector, will be presented in Section 2.4.1.

The remaining model types, which will be reviewed in the introduction, are only applicable in the case when  $\zeta(x, \xi)$  is a random variable. Thus we have  $\zeta(x, \xi) = t^T(\xi)x - h(\xi)$ , where the components of the  $n$ -dimensional random vector  $t(\xi)$  are the elements of the single row of  $T(\xi)$ .

Motivated by reliability theory, Prékopa [260] has developed a model which is built by utilizing the conditional expectation of the loss size. The quality measure is chosen as

$$\rho_{\text{cexp}}(\vartheta) := \mathbb{E}[-\vartheta \mid \vartheta < 0], \quad \vartheta \in \mathcal{L}_1^1.$$

Consequently,  $\rho_{\text{cexp}}(\vartheta)$  is the conditional expectation of the loss, given that a loss occurs. With the corresponding evaluation function, constraints of the form

$$\mathbb{E}_\xi[-\zeta(x, \xi) \mid \zeta(x, \xi) < 0] \leq \gamma$$

are included into the model, where the prescribed  $\gamma$  is a maximal tolerable conditional expected loss size. This model will be the subject of Section 2.4.2.

In the following discussion it will be convenient to consider positive values of  $\zeta(x, \xi)$  as losses and negative values as gains. A further idea to include the loss size and simultaneously also provide control on the probability of loss is utilizing quantiles. The first stochastic optimization model of this type has been proposed by Kataoka [178]. For a given  $0 < \alpha < 1$ , we utilize the following quality measure:

$$\rho_{\text{VaR}}^\alpha(\vartheta) := v(\vartheta, \alpha) := \min\{z \mid F_\vartheta(z) \geq \alpha\}, \quad \vartheta \in \mathcal{L}_1^0, \quad (2.23)$$

defined on the set of all random variables on  $(\Omega, \mathcal{F}, P)$ , and with  $F_\vartheta$  standing for the probability distribution function of  $\vartheta$ . In other words, for a given  $\alpha$ ,  $\rho_{\text{VaR}}^\alpha(\vartheta)$  is the left endpoint of the closed interval of  $\alpha$ -quantiles of  $\vartheta$ . This leads to the following evaluation function

$$v(x, \alpha) := \min\{z \mid \Psi(x, z) \geq \alpha\},$$

where  $\Psi(x, \cdot)$  denotes the probability distribution function of  $\zeta(x, \xi)$  for each fixed  $x$ , and with  $\alpha$  being a prescribed (high) probability level, for instance,  $\alpha = 0.95$ . This quality measure is widely used in the finance industry, it is called *Value at*

*Risk (VaR)* there. We will consider optimization problems involving  $v(x, \alpha)$  in Section 2.3. In general, it is quite difficult to build numerically tractable optimization models which are based on VaR. The main difficulty is that  $v(x, \alpha)$ , as a function of  $x$ , is not convex in general.

An interesting recent approach for building SLP models is due to Rockafellar and Uryasev [282]. The idea is to combine VaR and the conditional expectation approach. The following quality measure is chosen:

$$\rho_{\text{CVaR}}^\alpha(\vartheta) := \min_z \left[ z + \frac{1}{1-\alpha} \mathbb{E}[(\vartheta - z)^+] \right], \quad \vartheta \in \mathcal{L}_1^1.$$

The motivation for introducing this quality measure is twofold. On the one hand, utilizing a well-known fact from probability theory it can be shown that the solution set of the above minimization problem coincides with the set of  $\alpha$ -quantiles of the distribution of  $\vartheta$ . On the other hand, under the assumption that  $\vartheta$  has a continuous distribution function, we have

$$\rho_{\text{CVaR}}^\alpha(\vartheta) = \mathbb{E}[\vartheta \mid \vartheta \geq v(\vartheta, \alpha)], \quad \vartheta \in \mathcal{L}_1^1,$$

where  $v(\vartheta, \alpha)$  is the value at risk (VaR) (see (2.23)). This means that  $\rho_{\text{CVaR}}^\alpha(\vartheta)$  is the conditional expectation of the loss given that the loss exceeds VaR. The evaluation function

$$v_c(x, \alpha) := \rho_{\text{CVaR}}^\alpha(\zeta(x, \xi))$$

has nice convexity properties. Therefore, the prototype problems will involve inequality constraints of the form  $v_c(x, \alpha) \leq \gamma$  and being in the objective,  $v_c(x, \alpha)$  will be minimized. A further attractive feature is that, for finite discrete distributions, the optimization problems can be reduced to linear programming problems. A detailed discussion of this model class will be the subject of Section 2.4.3.

Finally we consider modeling approaches where  $\zeta(x, \xi)$  is interpreted as a deviation between  $t^T(\xi)x$  and  $h(\xi)$ , with the quality measures penalizing this deviation. Admittedly, most quality measures which have been introduced so far, can also be interpreted from the purely mathematical viewpoint as measuring deviation. Nevertheless, we have chosen to discuss those quality measures as a separate class, which correspond to the following modeling attitude: both  $t^T(\xi)x$  and  $h(\xi)$  represent important quantities in their own right, and the emphasis in modeling risk is on their deviation. Deviations are interpreted as risk and therefore the quality measure will be called a *risk measure* in this context. As a typical example let us mention portfolio optimization in finance, where  $t^T(\xi)x$  represents the random portfolio return and  $h(\xi)$  models some benchmark return. For this approach see, for instance, Elton et al. [85] and also Section 2.7.3.

Our first example of a deviation measure is the risk measure

$$\rho_Q(\vartheta) := \sqrt{\mathbb{E}[\vartheta^2]}, \quad \vartheta \in \mathcal{L}_1^2$$

defined on the linear space of random variables with finite second moment. The corresponding evaluation function is

$$Q(x) = \sqrt{\mathbb{E}_\xi[(t^T(\xi)x - h(\xi))^2]}.$$

As a second example we take the mean absolute deviation, with the risk measure

$$\rho_A(\vartheta) := \mathbb{E}[|\vartheta|], \quad \vartheta \in \mathcal{L}_1^1$$

and the evaluation function

$$A(x) = \mathbb{E}_\xi[|t^T(\xi)x - h(\xi)|].$$

Stochastic programming models, based on risk measures of this type, will be the subject of Section 2.5. Let us mention that stochastic optimization models in this class are by some authors also termed as *robust optimization problems*.

The basic question concerning the various quality measures is, how the stochastic optimization problems, based on these measures, behave from the numerical point of view. This will be the main subject of the present chapter.

From the point of view of efficient numerical solution, the most desirable property of a nonlinear optimization problem is that it should be a convex programming problem. Regarding the above-formulated prototype problems (2.7) and (2.9), in a strict sense these would count as convex programming problems under the assumption that  $V$  and  $V_i$  are concave functions.

For the subsequent discussion we will assume that in the objective function of (2.9) the additive linear term  $c^T x$  is missing, that is, we assume that  $c = 0$  holds. The reason for this assumption is that we will work with functions  $V$  having some generalized concavity properties. For such functions the addition of a linear term may destroy the generalized concavity property. Examples for this phenomenon will be presented later on in this section.

We will employ the following generalization of the notion of a convex programming problem: we consider the above-mentioned problems as convex programming problems, if the feasible domain is convex and if  $V(x)$  is a pseudo-concave function in (2.9). For general properties of optimization problems of this type see, for instance, Bazaraa and Shetty [11] and Avriel, Diewert, Schaible, and Zang [8].

We proceed with a short discussion concerning some generalizations of concave functions which will be utilized in this chapter.

**Definition 2.11.** Let  $f : C \rightarrow \mathbb{R}$  be a function defined over the convex set  $C$ .

- $f$  is called *quasi-concave*, if the inequality

$$f(\lambda x + (1 - \lambda)y) \geq \min\{f(x), f(y)\}$$

holds, for all  $x \in C$ ,  $y \in C$ , and  $\lambda \in [0, 1]$ .

- $f$  is called *quasi-convex*, if  $-f$  is quasi-concave.

Functions which are both quasi-convex and quasi-concave will be called *quasi-linear*. It is easy to see that  $f$  is quasi-concave if and only if the upper-level sets

$$\mathcal{U}_\gamma := \{x \mid f(x) \geq \gamma\} \quad (2.24)$$

are convex sets, for all  $\gamma \in \mathbb{R}$ . Thus, for ensuring the convexity of the feasible domain in (2.7), it will be sufficient to ensure that the function  $V$  is quasi-concave.

**Definition 2.12.** *Let  $f : C \rightarrow \mathbb{R}$  be a continuously differentiable function defined over an open convex set  $C$ .*

- *$f$  is called pseudo-concave, if the following implication*

$$\nabla^T f(x)(y - x) \leq 0 \implies f(y) \leq f(x)$$

*holds for all  $x \in C$  and  $y \in C$ .*

- *$f$  is called pseudo-convex, if  $-f$  pseudo-concave.*

The following facts are easy to check and are left as exercises for the reader: If  $f$  is a concave function, then it is quasi-concave and in the differentiable case it is also pseudo-concave. Pseudo-concave functions are also quasi-concave, for this assertion see e.g. Bazaraa and Shetty [11].

From our point of view, for maximization problems with quasi-concave restrictions (implying a convex feasible domain) and a pseudo-concave objective function, the most important properties are the following, see [11]:

- All local optimal solutions are global solutions.
- The Kuhn–Tucker optimality conditions are sufficient conditions of optimality.

Thus, in (2.9),  $V$  should be a pseudo-concave function. Note that requiring only quasi-concavity for  $V$ , results in general in non-convex optimization problems. Such problems may have local maxima which are not global.

A further remark concerns the quasi-concavity requirement for the constraint function  $V$  in (2.7). Although this way the convexity of the feasible domain is ensured, quasi-concavity is a rather weak property from the algorithmic point of view. One of the difficulties is that regularity conditions, which ensure the necessity of the Kuhn–Tucker conditions, are difficult to check in this case. From the algorithmic point of view it is much better, when besides the objective function, the constraint functions are pseudo-concave too. This implies, for instance, that the Slater-regularity can be utilized for enforcing the necessity of the Kuhn–Tucker conditions.

We will need the following fact concerning the pseudo-concavity of fractional functions:

**Proposition 2.32.** *Let  $f$  and  $g$  be continuously differentiable functions defined on  $\mathbb{R}^n$  and let  $C \subset \mathbb{R}^n$  be a convex set. We assume that  $f(x) \geq 0$  and  $g(x) > 0$  hold for all  $x \in C$ . If  $f$  is concave and  $g$  is convex then  $h(x) := \frac{f(x)}{g(x)}$  is pseudo-concave on  $C$ .*

*Proof:* Let  $x \in C$ ,  $y \in C$ , and assume that  $\nabla^T h(x)(y-x) \leq 0$  holds. By straightforward computation this implies

$$g(x)\nabla^T f(x)(y-x) - f(x)\nabla^T g(x)(y-x) \leq 0.$$

Utilizing the concavity of  $f$ , the convexity of  $g$ , the nonnegativity of  $f$ , and the positivity  $g$ , we get the inequality  $g(x)f(y) - f(x)g(y) \leq 0$  which immediately yields  $h(y) - h(x) \leq 0$ .  $\square$

Concerning transformations of pseudo-concave functions, the following fact will also be needed later on:

**Proposition 2.33.** *Let  $C$  be an open convex set and let  $g$  be a continuously differentiable pseudo-concave or pseudo-convex function, defined on  $C$ . Let  $f : \mathbb{R} \mapsto \mathbb{R}$  be a continuously differentiable, strictly monotonically increasing function, with  $f'(x) \neq 0$  for all  $x \in \mathbb{R}$ . Then  $h(x) := f(g(x))$  is pseudo-concave or pseudo-convex on  $C$ , respectively.*

*Proof:* For the gradient of  $h$  the relation  $\nabla h(x) = f'(g(x))\nabla g(x)$  obviously holds. We assume that  $g$  is pseudo-concave, the proof for the pseudo-convex case runs analogously. Let  $x \in C$ ,  $y \in C$ , and  $\nabla^T h(x)(y-x) \leq 0$ . Utilizing our assumptions, from this we get  $\nabla^T g(x)(y-x) \leq 0$ . The pseudo-concavity of  $g$  implies  $g(y) \leq g(x)$  and the monotonicity of  $f$  finally yields  $h(y) \leq h(x)$ .  $\square$

Unfortunately, the sum of a linear and a pseudo-concave function is not necessarily pseudo-concave. As an example take  $f_1(x) = -x$  and  $f_2(x) = x + x^3$ . It is easy to see that both functions are pseudo-concave, whereas their sum  $f_1(x) + f_2(x) = x^3$  is not pseudo-concave. As a multivariate example let us take the function  $f(x_1, x_2) = x_1 + x_1^3 + x_2 + x_2^3$  which is the sum of two pseudo-concave functions. The graph and the contour lines of this function are displayed in Figure 2.1. The function is clearly not quasi-concave, therefore it is not pseudo-concave, either.

A further important class of generalized concave functions consists of logarithmically concave (logconcave) functions.

**Definition 2.13.** *Let  $f : C \rightarrow \mathbb{R}$  be a nonnegative function defined over the convex set  $C$ .*

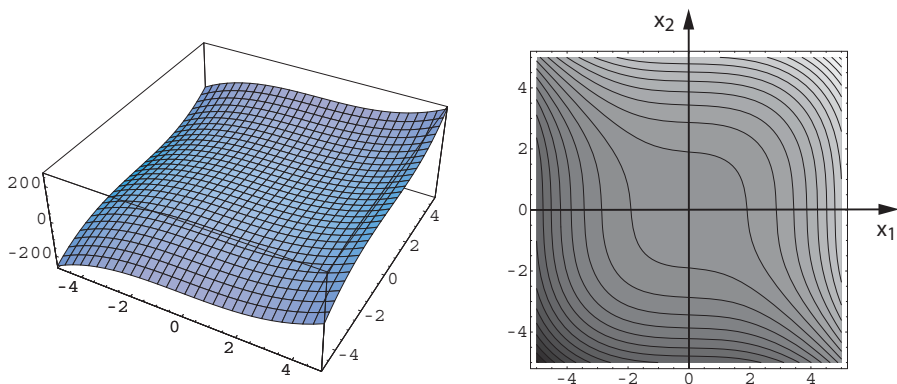
- *$f$  is called logarithmically concave or logconcave, if the inequality*

$$f(\lambda x + (1-\lambda)y) \geq [f(x)]^\lambda [f(y)]^{(1-\lambda)}$$

*holds, for all  $x \in C$ ,  $y \in C$ , and  $\lambda \in (0, 1)$ .*

- *$f$  is called logarithmically convex or logconvex, if the reverse inequality holds above.*

The definition immediately implies that for logconcave functions the set  $\mathcal{C}^+ := \{x \mid f(x) > 0, x \in C\}$  is convex. Observe, that the inequality in Definition 2.13. holds trivially, if either  $x \notin \mathcal{C}^+$  or  $y \notin \mathcal{C}^+$ . This leads to the following simple alternative characterization of logconcave functions:



**Fig. 2.1** The sum of two pseudo-concave functions needs not to be pseudo-concave. The picture shows the graph and the contour lines of the function  $f(x_1, x_2) = x_1 + x_1^3 + x_2 + x_2^3$ .

**Proposition 2.34.** *A nonnegative function  $f$  is logconcave over the convex set  $C$ , if and only if  $\mathcal{C}^+ = \{x \mid f(x) > 0, x \in C\}$  is a convex set and  $\log f$  is a concave function over  $\mathcal{C}^+$ .*

The next property involves products of logconcave functions. Let  $f_i, i = 1, \dots, r$ , be logconcave functions on a convex set  $C$  and let  $\mathcal{C}_i^+ := \{x \mid f_i(x) > 0, x \in C\}$  as before, for all  $i$ . Then the product  $f(x) = \prod_{i=1}^r f_i(x)$  is also logconcave on  $C$ . In fact, let us observe that

$$C^+ := \{x \mid f(x) > 0, x \in C\} = \bigcap_{i=1}^r C_i^+$$

holds. Thus  $C^+$  is a convex set and the assertion follows by considering  $\log f$  on  $C^+$ .

A further fact concerning logconcave functions, which will be needed later on, is the following. Let  $f$  be a logconcave function on  $\mathbb{R}^n$ . Then  $g(x) := f(x+y)$  is also logconcave on  $\mathbb{R}^n$  for any fixed  $y \in \mathbb{R}^n$ . Moreover,  $h(x, y) := f(x+y)$  is logconcave on  $\mathbb{R}^{2n}$ . In fact, for arbitrary  $u, v \in \mathbb{R}^n$  and  $\lambda \in (0, 1)$  we have  $g(\lambda u + (1-\lambda)v) = f(\lambda(u+y) + (1-\lambda)(v+y))$  from which the first assertion follows immediately. The second assertion follows also easily from the definition of logconcavity.

Considering logconvex functions, the definition implies the convexity of the set  $\mathcal{C}^0 := \{x \mid f(x) = 0, x \in C\}$ . Let  $\text{rint}C$  stand for the relative interior of  $C$  (see, for instance, Rockafellar [281]). It is easy to see, that  $\text{rint}C \cap \mathcal{C}^0 \neq \emptyset$  implies that  $\text{rint}C \subset \mathcal{C}^0$  holds. Thus, a logconvex function  $f$  for which  $\text{rint}C \cap \mathcal{C}^0 \neq \emptyset$  holds, can only have positive values at the (relative) boundary. Such functions are of no interest to us, therefore we will only consider positive logconvex functions. If  $f(x) > 0$  for all  $x \in C$ , then  $f$  is logconvex, if and only if  $\log f$  is convex. Finally let us remark that logconvex functions are also convex. This follows immediately from the inequality between the geometric and arithmetic means, see, for instance, Hardy et al. [132].

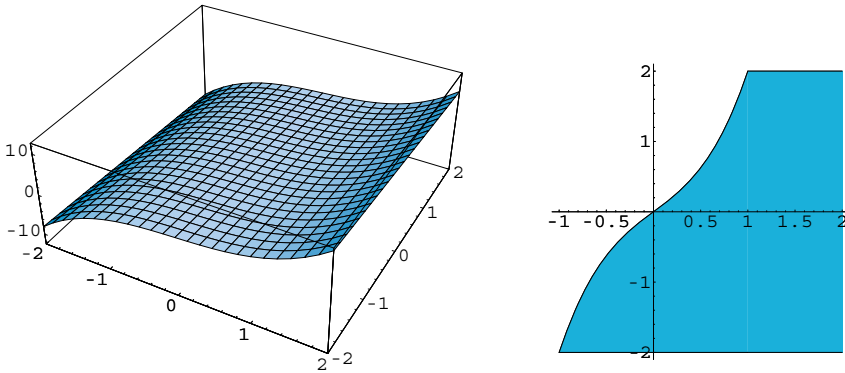
For further properties of logconcave and logconvex functions see, for instance, Kallberg and Ziemba [173] and Prékopa [266].

In the differentiable case, the class of strictly positive logconcave functions is a subset of the class of pseudo-concave functions:

**Proposition 2.35.** *Let  $f$  be a continuously differentiable, strictly positive, logconcave function over the open convex set  $C$ . Then  $f$  is pseudo-concave over  $C$ .*

*Proof:* Let  $x \in C$ ,  $y \in C$ ,  $\lambda \in [0, 1]$ , and assume that  $\nabla^T f(x)(y-x) \leq 0$  holds. This implies that  $\nabla^T \log f(x)(y-x) = \frac{1}{f(x)} \nabla^T f(x)(y-x) \leq 0$  also holds. However,  $\log f(x)$  being a concave function, it is also pseudo-concave, and consequently we have  $\log f(y) \leq \log f(x)$ , which implies the assertion immediately.  $\square$

Let us remark, that the notion of pseudo-concave functions can be extended to the non-differentiable case, see, for instance, [8]. We will not need this generalization in this book.



**Fig. 2.2** The graph of the function  $x + x^3 - z$  and the set  $\{(x, z) \mid x + x^3 - z \geq 0\}$ .

Finally, let us discuss a popular trick for equivalently reformulating the optimization problem (2.9) as follows:

$$\left. \begin{array}{ll} \max & c^T x + z \\ & V(x) - z \geq 0 \\ \text{s.t.} & x \in \mathcal{B}. \end{array} \right\} \quad (2.25)$$

This reformulation is used, for example, if we wish to apply cutting plane methods for solving (2.9). If  $V$  is a concave function, then (2.25) is obviously a convex programming problem. If, however,  $V$  is merely pseudo-concave, then this is in general not true. An example involving the pseudo-concave function  $x + x^3$  is displayed in Figure 2.2;  $x + x^3 - z$  is not quasi-concave and the feasible domain of the corre-



sponding problem (2.25) is a non-convex set. Thus, the reformulated problem (2.25) is in general much harder to solve than the problem in the original formulation.

Requiring the stronger property of logconcavity (cf. Proposition 2.35.) does not help, either. Take  $e^x$  as an example. This function is obviously logconcave, whereas  $e^x - z$  is a nonlinear convex function and the upper level set  $\{(x, z) \mid e^x - z \geq 0\}$  is a non-convex set. Thus,  $e^x - z$  is obviously not logconcave, in fact, it is not quasi-concave.

As already mentioned above, we will call our optimization problems (2.7) and (2.9) convex programming problems, if  $V$  is pseudo-concave in (2.9) and  $V$  is quasi-concave in (2.7), respectively. Whether or not our optimization problems are of the convex programming type, depends solely on (generalized) concavity properties of the function  $V$ .

## Exercises

**2.1.** Let  $f$  be a concave function defined on an open convex set  $C$ . Show that

- (a)  $f$  is quasi-concave;
- (b) if  $f$  is differentiable then it is pseudo-concave;
- (c) if  $f(x) > 0$  holds for all  $x \in C$  then  $f$  is logconcave.

**2.2.** For each of the following functions determine whether they are quasi-convex, quasi-concave, pseudo-convex, pseudo-concave, logconvex, or logconcave.

With  $C = \mathbb{R}$  let  $f_1(x) = e^x$  and  $f_2(x) = x^3$ . With  $C = \mathbb{R}^2$  let  $f_3(x_1, x_2) = e^{-x_1^2 - x_2^2}$  and let  $f_4$  be the indicator function of a set  $\mathcal{B}$ , with  $\mathcal{B}$  being a convex proper subset of  $\mathbb{R}^2$ . Formally,  $f_4(x) = \begin{cases} 1, & \text{if } x \in \mathcal{B}, \\ 0, & \text{if } x \notin \mathcal{B}. \end{cases}$

**2.3.** Let  $f$  and  $g$  be positive logconvex functions defined on the convex set  $C$ . Show that their sum  $h(x) = f(x) + g(x)$  is also logconvex over  $C$ .

## 2.2 Models involving probability functions

This section is devoted to pursuing the idea of using probability as a quality measure. We choose the following quality measure for evaluating random vectors

$$\rho_p(\vartheta) := \mathbb{P}(\vartheta \geq 0), \quad \vartheta \in \mathcal{L}_s^0, \quad (2.26)$$

which is defined on the set of all random vectors on  $(\Omega, \mathcal{F}, P)$ . The decision vector  $x$  will be evaluated by the corresponding evaluation function  $G(x) := \rho_p(\zeta(x, \xi)) := \mathbb{P}_\xi(\zeta(x, \xi) \geq 0)$ . The function  $G$  will be called a *probability function*. In a detailed form we have

$$G(x) := \mathbb{P}_\xi ( T(\xi)x - h(\xi) \geq 0 ). \quad (2.27)$$

Let  $x \in \mathbb{R}^n$  be fixed arbitrarily and let  $S(x) := \{z \in \mathbb{R}^r \mid T(z)x - h(z) \geq 0\}$ . Due to our assumptions,  $T(\cdot)$  and  $h(\cdot)$  are affine linear functions (see (2.2) on page 71). Consequently,  $S(x) \subset \mathbb{R}^r$  is a polyhedral set and

$$G(x) = \mathbb{P}(\xi \in S(x)) \quad (2.28)$$

holds.

The following prototype problems will be considered:

$$\left. \begin{array}{ll} \max & c^T x \\ \text{s.t.} & \mathbb{P}_\xi ( T(\xi)x - h(\xi) \geq 0 ) \geq \alpha \\ & x \in \mathcal{B} \end{array} \right\} \quad (2.29)$$

and

$$\left. \begin{array}{ll} \max & \mathbb{P}_\xi ( T(\xi)x - h(\xi) \geq 0 ) \\ \text{s.t.} & x \in \mathcal{B}, \end{array} \right\} \quad (2.30)$$

where  $\mathcal{B}$  is a polyhedral set given, for example, in the standard form

$$\mathcal{B} = \{x \mid Ax = b, l \leq x \leq u\}.$$

In this section we will assume throughout that  $\mathcal{B} \neq \emptyset$  holds and that  $\mathcal{B}$  is bounded.

Both optimization problems (2.29) and (2.30) are non-convex optimization problems in general. The emphasis in this section will be laid on identifying those subclasses, for which (2.29) and (2.30) belong to the class of convex optimization problems. We will throughout first consider the basic properties of the models above and will subsequently discuss the analogous results for the models with reversed direction of the inequality constraint and of optimization, respectively.

Notice that (2.30) is formulated without an additive linear term in the objective function. In the case, when the probability function is concave, the objective function in (2.30) would obviously remain concave with an additive linear term. However, in general, we will only be able to ensure some generalized concavity properties of probability functions, which are usually lost when adding a linear function to them.

As already mentioned above, the function  $G$  will be called a probability function. The constraint involving a probability function in (2.29) is called a *chance-constraint* or a *probabilistic constraint*. For constraints involving probability functions the following terminology will be used. In the case of  $s = 1$  the constraint will be called *separate*, whereas in the case when  $s > 1$  is permitted, the term *joint constraint* will be used. In this sense, *joint constraint* stands for the general case, which specializes to a *separate constraint* if  $s = 1$  holds. The corresponding probability functions will be called joint and separate probability functions, respectively. This terminology has its roots in modeling. Let us consider a joint probability constraint

$$\mathbb{P}_\xi (t_i^T(\xi)x \geq h_i(\xi), i = 1, \dots, s) \geq \alpha,$$

where the components of  $t_i(\xi)$  are the elements of the  $i^{\text{th}}$  row of  $T(\xi)$  and let us assume that  $s > 1$  holds. In this constraint, the underlying event has the following interpretation: a system of random inequalities holds, meaning that all of the inequalities hold simultaneously (they hold *jointly*). Depending on the modeling situation, we may wish to consider separately for  $i = 1, \dots, s$  the events that the  $i^{\text{th}}$  random inequality  $t_i^T(\xi)x \geq h_i(\xi)$  holds. In this case, the joint constraint above is split into  $s$  separate probability constraints, where the probability levels on the right-hand-side can now be chosen differently for different rows:

$$\mathbb{P}_\xi(t_i^T(\xi)x \geq h_i(\xi)) \geq \alpha_i, \quad i = 1, \dots, s.$$

Let us make a further remark concerning terminology. In the literature, model (2.29) is called either *chance constrained* or alternatively, *probabilistic constrained* model. Both *chance* and *probabilistic* have a very general meaning, including virtually all aspects of randomness. None of them describes with sufficient accuracy the fact that we are dealing with constraints and objective functions which are defined via probabilities. In order to contrast models involving probability functions with other SLP models based on different quality measures, we use a terminology, which explicitly refers to probability. For this reason, we call  $G$  a probability function. This terminology has been coined by Uryasev, see, for instance, [331]. With our notations, a probability function in [331] is defined as a function of the following type:

$$\mathbb{P}_\xi(f(x, \xi) \geq 0),$$

where  $f(x, \cdot)$  is Borel-measurable for all  $x$ . Our case fits this scheme by choosing  $f(x, \xi) = T(\xi)x - h(\xi)$ . In accordance with this, models like (2.29) and (2.30) will be generally called *SLP models with probability functions*.

Next we discuss the reformulation of the constraint  $G(x) \geq \alpha$ , as an equivalent constraint with reversed inequality. We have

$$\begin{aligned} \mathbb{P}_\xi(\zeta(x, \xi) \geq 0) \geq \alpha &\iff \mathbb{P}_\xi([\min_{1 \leq i \leq s} \zeta_i(x, \xi)] \geq 0) \geq \alpha \\ &\iff \mathbb{P}_\xi([\min_{1 \leq i \leq s} \zeta_i(x, \xi)] < 0) \leq 1 - \alpha \\ &\iff \mathbb{P}_\xi([\max_{1 \leq i \leq s} (\zeta_i(x, \xi)^-)] > 0) \leq 1 - \alpha, \end{aligned} \tag{2.31}$$

where for any real number  $z$ ,  $z^- := \max\{0, -z\}$  denotes the negative part of  $z$ . Note that, in comparison with the original probability function  $G(x)$ , the probability function on the left-hand-side of the equivalent reversed inequality is much more difficult to handle numerically. On the one hand, the underlying event in the probability function involves a strict inequality. On the other hand, for computing this probability function for a fixed  $x$ , the probability measure of the region  $\mathbb{R}^r \setminus S(x)$  is to be computed, which is the complement of a polyhedral set and thus it is non-convex in general (cf. (2.28)). In the special case  $s = 1$  the situation is much simpler:  $S(x)$  is a half-space and thus  $\mathbb{R}^r \setminus S(x)$  becomes an open half-space. (2.31) reduces to

$$\begin{aligned} \mathbb{P}_\xi(\zeta(x, \xi) \geq 0) \geq \alpha &\iff \mathbb{P}_\xi(\zeta(x, \xi)^- > 0) \leq 1 - \alpha \\ &\iff \mathbb{P}_\xi(t^T(\xi)x - h(\xi) < 0) \leq 1 - \alpha, \end{aligned} \quad (2.32)$$

where the components of  $t(\xi)$  are the elements of the single row of  $T(\xi)$ . This is the straightforward way for reversing a separate probability constraint. We still have a strict inequality which can be replaced by an inequality involving “ $\leq$ ”, if the probability distribution function of  $\zeta(x, \xi)$  is continuous.

We will also need a reformulation of (2.31) in expectation terms:

$$\mathbb{P}_\xi(\zeta(x, \xi) \geq 0) \geq \alpha \iff \mathbb{E}_\xi[\chi(\max_{1 \leq i \leq s} \zeta_i(x, \xi)^-)] \leq 1 - \alpha, \quad (2.33)$$

where  $\chi$  is the following indicator function

$$\chi(z) = \begin{cases} 0 & \text{if } z \leq 0, \\ 1 & \text{if } z > 0. \end{cases}$$

For the set of vectors which are feasible with respect to the probability constraint, we introduce the notation

$$\mathcal{B}(\alpha) = \{x \mid G(x) \geq \alpha\} \quad (2.34)$$

and for the sake of easy reference we formulate our prototype problems (2.29) and (2.30) also in terms of the probability function  $G$  as follows:

$$\left. \begin{array}{ll} \max & c^T x \\ \text{s.t.} & G(x) \geq \alpha \\ & x \in \mathcal{B} \end{array} \right\} \quad (2.35)$$

and

$$\left. \begin{array}{ll} \max & G(x) \\ \text{s.t.} & x \in \mathcal{B}. \end{array} \right\} \quad (2.36)$$

**Remark.** Let us consider the case, when one of the rows of the matrix  $(T(\xi), h(\xi))$  is constant almost surely, for instance, it is deterministic. Denoting by  $t_i(\xi)$  the random vector with its components being the elements of the  $i^{\text{th}}$  row of  $T(\xi)$ , we assume without loss of generality that  $(t_1^T(\xi), h_1(\xi)) = (t^T, h)$  a.s. holds, where  $t \in \mathbb{R}^n$  and  $h \in \mathbb{R}$  are deterministic. In this case

$$\mathcal{B}(\alpha) = \{x \mid \mathbb{P}_\xi(t_i(\xi)x \geq h_i(\xi), i = 2, \dots, s) \geq \alpha\} \cap \{x \mid t^T x \geq h\}$$

holds. This implies, that  $\mathcal{B}(\alpha) \cap \mathcal{B}$  remains unchanged if  $G$  and  $\mathcal{B}$  are redefined as follows:

$$\begin{aligned} G(x) &:= \mathbb{P}_\xi(t_i(\xi)x \geq h_i(\xi), i = 2, \dots, s) \\ \mathcal{B} &:= \mathcal{B} \cap \{x \mid t^T x \geq h\}. \end{aligned}$$

The meaning is the following: essentially deterministic inequalities within a probability constraint can be removed from this constraint, by appending them to the set

of deterministic constraints.  $\square$

As already discussed in the introductory section 2.1, our optimization problems will be considered as convex programming problems, if  $G$  is pseudo-concave in (2.36), and if it is quasi-concave in (2.35). It may happen, however, that  $G$  is not a quasi-concave function but nevertheless (2.35) is a convex programming problem. The point is this. As we have discussed in the introduction to this chapter on page 83, a function is quasi-concave if and only if all upper level sets are convex. The domain  $\mathcal{B}(\alpha)$  defined in (2.34) is clearly an upper level set corresponding to level  $\alpha$ . The convexity of the feasible domain of (2.35) just means that this specific level set is convex. It will turn out that, for some model classes and probability distributions,  $\mathcal{B}(\alpha)$  becomes convex for  $\alpha$  large enough. In summary: whether or not (2.35) is a convex programming problem, may also depend on the prescribed probability level  $\alpha$ .

### 2.2.1 Basic properties

The purpose of this section is to present some general results which hold without any assumptions concerning the probability distribution of  $\xi$ .

We consider the probability function

$$G(x) = \mathbb{P}_{\xi}(T(\xi)x \geq h(\xi))$$

as well as the constraint involving this probability function

$$G(x) \geq \alpha. \tag{2.37}$$

This constraint requires, that for a feasible  $x$  the event

$$S(x) := \{\xi \mid T(\xi)x \geq h(\xi)\} \in \mathbb{B}^r$$

should belong to the set of events  $\mathcal{G}_{\alpha}$  having probability measure of at least  $\alpha$

$$\mathcal{G}_{\alpha} = \{A \in \mathbb{B}^r \mid \mathbb{P}_{\xi}(A) \geq \alpha\}.$$

For the feasible set, determined by (2.37) and denoted by

$$\mathcal{B}(\alpha) = \{x \mid G(x) \geq \alpha\} = \{x \mid S(x) \in \mathcal{G}_{\alpha}\},$$

the following representation holds obviously:

$$\begin{aligned}
\mathcal{B}(\alpha) &= \bigcup_{A \in \mathcal{G}_\alpha} \{x \mid S(x) = A\} \\
&= \bigcup_{A \in \mathcal{G}_\alpha} \{x \mid \forall \xi \in A : T(\xi)x \geq h(\xi)\} \\
&= \bigcup_{A \in \mathcal{G}_\alpha} \bigcap_{\xi \in A} \{x \mid T(\xi)x \geq h(\xi)\}.
\end{aligned} \tag{2.38}$$

Both from the theoretical point of view concerning the existence of optimal solutions and from the standpoint of numerical solution it is an important question whether  $\mathcal{B}(\alpha)$  is a closed set. The answer is affirmative:

**Theorem 2.1.** *The set  $\mathcal{B}(\alpha)$  is closed.*

*Proof:* For a proof see Kall and Wallace [172], Proposition 1.7.  $\square$

Without any assumptions on the probability distribution of  $\xi$ , the sole available result concerning the convexity of  $\mathcal{B}(\alpha)$  is the following:

**Theorem 2.2.** *Kall ([154]).  $\mathcal{B}(\alpha)$  is convex for  $\alpha = 0$  and  $\alpha = 1$ .*

*Proof:* For  $\alpha = 0$  we clearly have  $\mathcal{G}_\alpha = \mathbb{B}^r$  and consequently  $\mathcal{B}(\alpha) = \mathbb{R}^n$  holds. For the case  $\alpha = 1$  we first observe that  $A \in \mathcal{G}_1$  and  $B \in \mathcal{G}_1$  imply  $A \cap B \in \mathcal{G}_1$  (consider the complement of  $A \cap B$ ). Now let  $x \in \mathcal{B}(1)$ ,  $y \in \mathcal{B}(1)$ ,  $\lambda \in [0, 1]$ , and  $z = \lambda x + (1 - \lambda)y$ . Then we have  $S(x) \in \mathcal{G}_1$  and  $S(y) \in \mathcal{G}_1$  and consequently  $S(x) \cap S(y) \in \mathcal{G}_1$ . For arbitrary fixed  $\xi \in \mathbb{R}^r$ , the inequalities  $T(\xi)x \geq h(\xi)$  and  $T(\xi)y \geq h(\xi)$  obviously imply the inequality  $T(\xi)z \geq h(\xi)$ . Thus  $S(x) \cap S(y) \subset S(z)$  holds, implying  $S(z) \in \mathcal{G}_1$ .  $\square$

In the case of  $\alpha = 0$  the probability constraint is clearly redundant. If  $\alpha = 1$ , then the solution of (2.35) can be interpreted as a “fat solution”, in a probabilistic sense.

Finally let us discuss the reverse inequality  $G(x) \leq \beta$ . We consider now

$$\mathcal{H}_\beta = \{A \in \mathbb{B}^r \mid \mathbb{P}_\xi(A) \leq \beta\}$$

and denoting the feasible set in this case also by  $\mathcal{B}(\beta)$  we have

$$\mathcal{B}(\beta) = \{x \mid G(x) \leq \beta\} = \{x \mid S(x) \in \mathcal{H}_\beta\}.$$

Analogously as above, we get the following representation:

$$\mathcal{B}(\beta) = \bigcup_{A \in \mathcal{H}_\beta} \bigcap_{\xi \in A} \{x \mid T(\xi)x \geq h(\xi)\}. \tag{2.39}$$

Considering the analogous assertion to Theorem 2.2.,  $\mathcal{B}(1) = \mathbb{R}^n$  is obviously convex and the probability constraint is redundant.  $\mathcal{B}(0)$  is in general not convex, though. To see this, let us consider the following example with  $x \in \mathbb{R}^1$ ,  $G(x) = \mathbb{P}(x \geq \xi_1, -x \geq \xi_2)$  where  $\xi$  has the singular distribution  $\xi_1 \equiv -1$ ,  $\xi_2 \equiv -1$ . We have  $\mathcal{B}(0) = (-\infty, -1) \cup (1, \infty)$  which is obviously not convex.

### 2.2.2 Finite discrete distribution

We consider the case, when  $\xi$  has a finite discrete distribution, given by a realization tableau

$$\begin{pmatrix} p_1 & \dots & p_N \\ \hat{\xi}^1 & \dots & \hat{\xi}^N \end{pmatrix} \quad (2.40)$$

with  $p_i > 0 \forall i$  and  $\sum_{i=1}^N p_i = 1$ .

The discussion will be focused on the model (2.35), formulated as follows

$$\left. \begin{array}{ll} \max & c^T x \\ \text{s.t.} & x \in \mathcal{B}(\alpha) \cap \mathcal{B} \end{array} \right\} \quad (2.41)$$

with  $\mathcal{B}(\alpha) = \{x \mid G(x) \geq \alpha\}$ .

In the discretely distributed case the representation (2.38) on page 92 specializes as follows. Let  $I = \{1, \dots, N\}$ , then we have

$$\mathcal{B}(\alpha) = \bigcup_{\substack{J \subset I \\ \sum_{j \in J} p_j \geq \alpha}} \bigcap_{j \in J} \{x \mid T(\hat{\xi}^j)x \geq h(\hat{\xi}^j)\}. \quad (2.42)$$

For the separate realizations of  $\xi$  let us introduce the notation

$$K_j = \{x \mid T(\hat{\xi}^j)x \geq h(\hat{\xi}^j)\}, \quad j = 1, \dots, N.$$

These sets are clearly convex polyhedral sets. Employing this notation, the representation above can be written in the form

$$\mathcal{B}(\alpha) = \bigcup_{\substack{J \subset I \\ \sum_{j \in J} p_j \geq \alpha}} \bigcap_{j \in J} K_j. \quad (2.43)$$

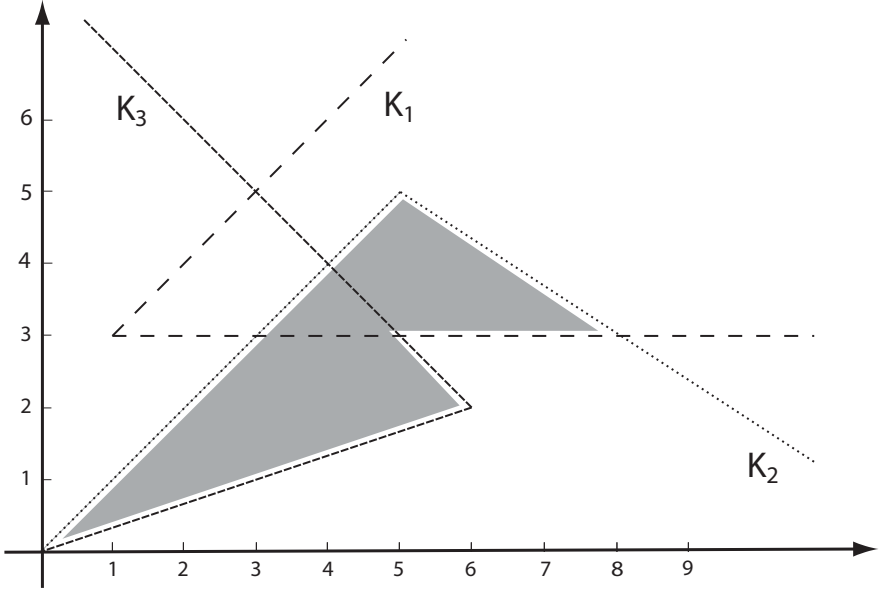
Figure 2.3 shows the following example from Kall [154]:

$$\begin{aligned} K_1 &= \{x \in \mathbb{R}^2 \mid x_1 - x_2 \geq -2, x_2 \geq 3\}, \\ K_2 &= \{x \in \mathbb{R}^2 \mid x_1 - x_2 \geq 0, 2x_1 + 3x_2 \leq 25\}, \\ K_3 &= \{x \in \mathbb{R}^2 \mid x_1 + x_2 \leq 8, -x_1 + 3x_2 \geq 0\} \end{aligned}$$

with corresponding probabilities of realizations  $p_1 = \frac{1}{4}$ ,  $p_2 = \frac{1}{2}$ , and  $p_3 = \frac{1}{4}$ . The probability level in the probability constraint is  $\alpha = \frac{3}{4}$ . The feasible domain is the shaded region in the figure, which is obviously non-convex. The following representation holds:  $\mathcal{B}(\alpha) = [K_1 \cap K_2] \cup [K_2 \cap K_3]$ .

A necessary condition for  $\mathcal{B}(\alpha) \cap \mathcal{B} \neq \emptyset$  is the following. With the notation  $I_0 = \{i, 1 \leq i \leq N \mid K_i \cap \mathcal{B} = \emptyset\}$ ,  $\mathcal{B}(\alpha) \cap \mathcal{B} \neq \emptyset$  obviously implies that

$$\sum_{i \notin I_0} p_i \geq \alpha$$



**Fig. 2.3** Feasible domain of an SLP problem with joint probability constraint and finite discrete distribution.

must hold, otherwise each of the intersections in (2.43) would involve at least one  $j \in I_0$ , which would lead after intersecting  $\mathcal{B}(\alpha)$  with  $\mathcal{B}$  to a union of empty sets.

From (2.43) it is immediately clear, that our optimization problem (2.41) involves maximizing a linear function over a *union* of convex polyhedral sets. Thus, in general, the optimization problems do not belong to the class of convex optimization problems. This type of problems is called *disjunctive programming problem*, see, for instance, Nemhauser and Wolsey [241].

Utilizing the usual transformation of disjunctive programming, an equivalent mixed-integer formulation of (2.41) is the following (Raïke [274]):

$$\left. \begin{array}{ll}
 \min & c^T x \\
 \text{s.t.} & T(\hat{\xi}^k)x + M \cdot (1 - z_k) \mathbb{1} \geq h(\hat{\xi}^k), \quad k = 1, \dots, N \\
 & \sum_{i=1}^N p_i z_i \geq \alpha \\
 & z_k \in \{0, 1\}, \quad k = 1, \dots, N \\
 & x \in \mathcal{B},
 \end{array} \right\} \quad (2.44)$$



where binary variables  $z_j$  have been introduced,  $M$  is a “big enough” constant and  $\mathbf{1}^T = (1, \dots, 1)$ .  $M$  is chosen in such a way, that  $M \geq h(\hat{\xi}^k) - T(\hat{\xi}^k)x$  holds,  $\forall x \in \mathcal{B}, k = 1, \dots, N$ . Under our assumptions ( $\mathcal{B} \neq \emptyset$ ,  $\mathcal{B}$  bounded), such an  $M$  can be computed, for instance, by solving the following linear programming problems for  $k \notin I_0$ :

$$M_k = \max\{\gamma \mid \gamma + T(\hat{\xi}^k)x \geq h(\hat{\xi}^k), x \in \mathcal{B}\}$$

and setting  $M = \max_{k \notin I_0} M_k$ .

For the case when only the right-hand-side is stochastic, further equivalent formulations as mixed-integer linear programming problems can be found in Prékopa [266].

There are some special cases, where the union in (2.42) amounts in a single convex polyhedral set.

**Theorem 2.3.** *Marti 1971 [222]. Let  $p_{i_0} = \min_{i \in I} p_i$ . Then  $\mathcal{B}(\alpha)$  is convex for  $\alpha > 1 - p_{i_0}$ .*

*Proof:* For the proof see Kall [154]. □

Notice that  $\alpha > 1 - p_{i_0}$  implies that  $\mathcal{B}(\alpha) = \mathcal{B}(1)$  holds. Consequently, the constraint involving a probability function (2.37) can be replaced by the system of linear inequalities

$$T(\hat{\xi}^i)x \geq h(\hat{\xi}^i), i = 1, \dots, N. \quad (2.45)$$

Requiring that the inequalities should hold for all realizations, results in a “fat solution”.

The result can be sharpened in a further special case:

**Theorem 2.4.** *Kall 1976 [154]. Let  $p_{i_0} = \min_{i \in I} p_i$  and assume that  $p_{i_0}$  is uniquely determined. Let  $p_{i_1} = \min_{i \in I \setminus \{i_0\}} p_i$ . Then  $\mathcal{B}(\alpha)$  is convex for  $\alpha > 1 - p_{i_1}$ .*

*Proof:* For the proof see Kall [154]. □

### 2.2.3 Separate probability functions

This section is devoted to discussing stochastic programming models which involve separate probability functions. The general prototype formulation of such problems has the same form as (2.29) and (2.30) with  $\zeta(x, \xi)$  now being a random variable ( $s = 1$ ). To emphasize one of the typical sources of such problems, we give a formulation for a random vector  $\zeta(x, \xi)$  where the evaluation function has been applied component-wise:

$$\left. \begin{array}{l} \max c^T x \\ \text{s.t. } \mathbb{P}_\xi( t_k^T(\xi)x \geq h_k(\xi) ) \geq \alpha_k, \quad k = 1, \dots, s \\ x \in \mathcal{B} \end{array} \right\} \quad (2.46)$$

and

$$\left. \begin{array}{l} \max \mathbb{P}_\xi( t^T(\xi)x \geq h(\xi) ) \\ \text{s.t. } x \in \mathcal{B} \end{array} \right\} \quad (2.47)$$

where the components of the  $n$ -dimensional random vector  $t_k(\xi)$  are the elements of the  $k^{\text{th}}$  row of  $T(\xi)$ ,  $\forall k$ ;  $t(\xi)$  is an  $n$ -dimensional random vector and  $h_k(\xi)$ ,  $h(\xi)$  are random variables  $\forall k$ . The term *separate* means, as we have discussed previously, that each of the probability functions appearing in the model formulations involves a single random inequality.

For the discussions regarding convexity of the feasible domain, it is clearly sufficient to consider a single separate probability function:

$$G(x) = \mathbb{P}_\xi(x \mid t(\xi)^T x \geq h(\xi)).$$

For the sake of simplicity we introduce the notation  $\eta := t(\xi)$  and replace the right-hand-side  $h(\xi)$  by  $\xi$ , because only the probability distribution of  $(t(\xi)^T, h(\xi))$  counts anyway. Thus the probability function has the following form:

$$G(x) = \mathbb{P}_\xi(x \mid \eta^T x - \xi \geq 0).$$

With our notation, the definition of  $\zeta(x, \xi)$  on page 71 takes the form

$$\zeta(x, \eta, \xi) := \eta^T x - \xi.$$

Note that  $\zeta(x, \eta, \xi)$  is now a random variable.

The goal of this section is to identify subclasses of SLP models with separate probability functions, which lead to convex programming problems. We will also give equivalent formulations for these models in algebraic terms, which provide the basis for the numerical solution of the problems. It will turn out for this class of models that both type of constraints  $G(x) \geq \alpha$  and  $G(x) \leq \beta$  can lead, under appropriate assumptions, to convex optimization problems.

We will proceed as follows. Next we will discuss the special case when only the right-hand-side is stochastic. This will be followed by considering the case when  $(\eta, \xi)$  has a multivariate normal distribution. Next the results will be generalized to the class of stable distributions. Finally we discuss a distribution-free approach.

Considering other distributions, we mention that in the case when the components of  $(\eta, \xi)$  are independent and have exponential distributions, Biswal et al. [29] have presented an equivalent algebraic formulation as an NLP problem.

### Only the right-hand-side is stochastic

We assume that  $\eta \equiv t$  holds, with  $t$  being deterministic. In this case the probability function has the form

$$G(x) = \mathbb{P}_\xi(x \mid t^T x \geq \xi).$$

For the case of reverse random inequalities  $t^T x \leq \xi$  we just consider the probability function corresponding to  $(-t, -\xi)$ . Denoting the probability distribution function of the random variable  $\xi$  by  $F_\xi$ , we have

$$G(x) = F_\xi(t^T x).$$

The probability distribution function of a random variable being monotonically increasing, it is both quasi-convex and quasi-concave (it is quasi-linear). It is easy to see that substituting a linear function into a quasi-convex function results in a quasi-convex function, the same being true in the quasi-concave case. Consequently,  $G(x)$  is both quasi-convex and quasi concave which immediately implies that both  $\{x \mid G(x) \geq \alpha\}$  and  $\{x \mid G(x) \leq \beta\}$  are convex sets. From the algorithmic point of view, however, it is desirable to obtain an explicit representation in terms of inequalities involving algebraic functions. This is easy to achieve in our case.

Considering first the constraint  $G(x) \geq \alpha$ , this is obviously equivalent to a linear constraint:

$$\mathbb{P}_\xi(x \mid t^T x \geq \xi) \geq \alpha \iff F_\xi(t^T x) \geq \alpha \iff t^T x \geq Q_\xi^-(\alpha),$$

where  $Q_\xi^-(\alpha)$  denotes the left end-point of the closed interval of  $\alpha$ -quantiles of  $F_\xi$  (for properties of quantiles see, for instance, Cramér [47]).

Turning our attention to the reverse constraint  $G(x) \leq \beta$  we observe that this can be written as  $F_\xi(t^T x) \leq \beta$ . Assuming that  $F_\xi$  is continuous (for instance,  $\xi$  has a continuous distribution), we obtain again an equivalent linear inequality

$$\mathbb{P}_\xi(x \mid t^T x \geq \xi) \leq \beta \iff F_\xi(t^T x) \leq \beta \iff t^T x \leq Q_\xi^+(\beta),$$

with  $Q_\xi^+(\beta)$  denoting the right end-point of the interval of  $\beta$ -quantiles of  $F_\xi$ .

For arbitrary distributions, the equivalent reformulation should be set up with care. If  $F_\xi$  is continuous at the point  $Q_\xi^+(\beta)$ , then the above formulation holds. If, however,  $F_\xi$  is discontinuous at  $Q_\xi^+(\beta)$ , then the equivalent formulation is the following

$$\mathbb{P}_\xi(x \mid t^T x \geq \xi) \leq \beta \iff F_\xi(t^T x) \leq \beta \iff t^T x < Q_\xi^+(\beta),$$

with a strict linear inequality implying the numerically unpleasant feature that the set  $\{x \mid G(x) \leq \beta\}$  is an open half-space. This aspect reflects an asymmetry between the two setups  $G(x) \geq \alpha$  and  $G(x) \leq \beta$  of the constraints.

Having, for instance, a finite discrete distribution for  $\xi$ , the theoretically correct reformulation may consist of the strict inequality above. From the modeling point of view this is usually not a real problem: the unfavorable event (loss) can mostly be formulated as a strict inequality  $\mathbb{P}_\xi(t^T x < \xi)$  and thus we get

$$\begin{aligned} \mathbb{P}_\xi(t^T x < \xi) \leq \beta &\iff 1 - \mathbb{P}_\xi(t^T x \geq \xi) \leq \beta \\ &\iff F(t^T x) \geq 1 - \beta \iff t^T x \geq Q_\xi^-(1 - \beta), \end{aligned}$$

that means, we obtain an equivalent linear constraint.

For discussing the situation concerning the objective function, we consider the problem (2.36) which in our case has the form

$$\left. \begin{array}{l} \max F_\xi(t^T x) \\ \text{s.t. } x \in \mathcal{B}. \end{array} \right\} \quad (2.48)$$

This is a linearly constrained nonlinear programming problem. Let us associate with (2.48) the following linear programming problem:

$$\left. \begin{array}{l} \max t^T x \\ \text{s.t. } x \in \mathcal{B}. \end{array} \right\} \quad (2.49)$$

If  $F_\xi$  is strictly monotone, then (2.48) and (2.49) are clearly equivalent. In the general case, some care is needed. Provided that (2.49) has an optimal solution, this will be an optimal solution also for (2.48). Under our assumptions ( $\mathcal{B} \neq \emptyset$ ,  $\mathcal{B}$  bounded) this is always the case. For an unbounded polyhedral set  $\mathcal{B}$  it may happen, however, that (2.49) has an unbounded objective over  $\mathcal{B}$ , whereas (2.48) has an optimal solution.

Analogous comments apply in the case when in (2.48) the objective is minimized.

## Multivariate normal distribution

In this section we discuss the case, when  $(\eta^T, \xi)^T$  has a joint multivariate normal distribution. For excluding the case already discussed in the previous section, we assume that  $\eta$  is stochastic, that means, that  $\exists d \in \mathbb{R}^n : \eta = d$  a.s.

**Definition 2.14.** *See, for example, Tong [328]. The  $r$ -dimensional random vector  $\zeta$  has a multivariate normal distribution, if there exist an  $(r \times s)$  matrix  $B$  and  $\mu \in \mathbb{R}^r$ , such that*

$$\zeta = B\tilde{\zeta} + \mu \quad (2.50)$$

*holds, where  $\tilde{\zeta}$  is an  $s$ -dimensional random vector with  $\tilde{\zeta}_i$  being stochastically independent and having a standard normal distribution,  $\forall i$ .*

Note that this definition allows for deterministic components of  $\zeta$ : if the  $i^{\text{th}}$  row of  $B$  is zero then we have  $\zeta_i \equiv \mu_i$ . From the definition immediately follows that

- $\mathbb{E}[\zeta] = \mu$  and
- $\Sigma = BB^T$ , where  $\Sigma$  denotes the covariance matrix of  $\zeta$

hold.

$\Sigma$  is clearly a symmetric positive semidefinite matrix. The multivariate normal distribution is called *non-degenerate*, if  $\Sigma$  is positive definite. This is the case if and only if  $B$  has full row rank. Otherwise the distribution is called *degenerate* or *singular*.

The multivariate normal distribution is uniquely determined by the expected-value vector  $\mu$  and the covariance matrix  $\Sigma$ , see, for instance, Tong [328]. We will use the notation  $\zeta \sim \mathcal{N}(\mu, \Sigma)$ , meaning that the random vector  $\zeta$  has a normal distribution with expected value vector  $\mu$  and covariance matrix  $\Sigma$ .

If the multivariate normal distribution is *non-degenerate*, then it is absolutely continuous w.r. to the Lebesgue-measure on  $\mathbb{R}^r$ , having the probability density function

$$f(y) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(y-\mu)^T \Sigma^{-1} (y-\mu)} \quad (2.51)$$

where  $|\Sigma|$  denotes the determinant of  $\Sigma$ .

Let  $R$  be the *correlation matrix* of  $\zeta$ , defined as

$$R_{i,j} = \frac{\Sigma_{ij}}{\sigma_i \sigma_j}, \quad \forall i, j$$

where  $\sigma_i$  and  $\sigma_j$  denote the standard deviations of  $\zeta_i$  and  $\zeta_j$ , respectively. The non-degenerate multivariate normal distribution is called *standard multivariate normal distribution*, if the expected value vector is the zero-vector and the standard deviation of the components of  $\zeta$  is 1. It is defined by the following density function

$$\varphi(y; R) = \frac{1}{(2\pi)^{\frac{n}{2}} |R|^{\frac{1}{2}}} e^{-\frac{1}{2} y^T R^{-1} y}. \quad (2.52)$$

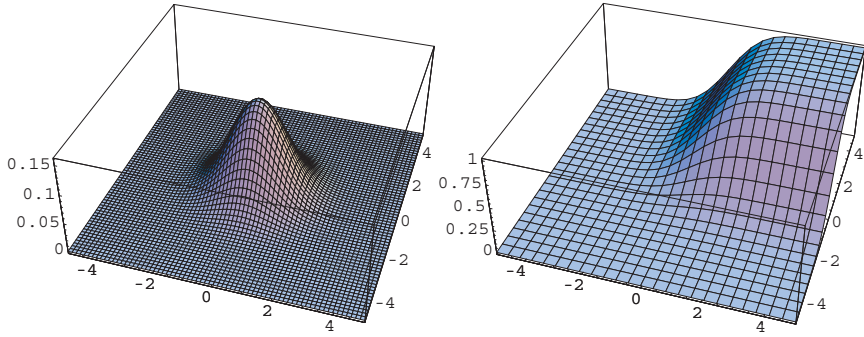
The corresponding distribution function will be denoted by  $\Phi(y; R)$ . In the univariate case we drop  $R$  in the notation;  $\varphi$  stands for the density function of the standard normal distribution, that means, we have

$$\varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

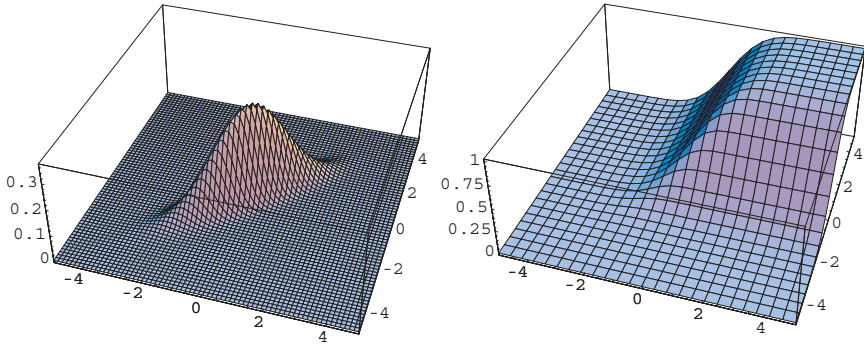
and the corresponding distribution function will be denoted by  $\Phi$ .

Figure 2.4 shows the density- and distribution functions of the bivariate normal distribution with correlation  $r = 0$ . In Figure 2.5 these functions are displayed for the case  $r = 0.9$ .

Having a symmetric positive semidefinite matrix  $\Sigma$  and vector  $\mu$  as primary data, a lower-triangular matrix  $B$  for relation (2.50) can be computed by the Cholesky-factorization for symmetric positive semidefinite matrices, see, for instance, Golub and Van Loan [127].



**Fig. 2.4** The bivariate normal distribution function with correlation  $r = 0$ .



**Fig. 2.5** The bivariate normal distribution function with correlation  $r = 0.9$ .

From the definition it follows immediately, that any affine linear transformation of a random vector with a multivariate normal distribution has again a multivariate normal distribution.

Assume now, that the  $(n + 1)$ -dimensional random vector  $\zeta^T = (\eta^T, \xi)^T$  has a multivariate normal distribution:

$$\zeta = \begin{pmatrix} \eta \\ \xi \end{pmatrix} = \begin{pmatrix} D \\ d^T \end{pmatrix} \cdot \tilde{\zeta} + \begin{pmatrix} \mu \\ \mu_{n+1} \end{pmatrix} \quad (2.53)$$

where  $D$  is an  $(n \times s)$  matrix,  $d \in \mathbb{R}^s$ ,  $\mu \in \mathbb{R}^n$ . We get

$$\begin{aligned} \zeta(x, \eta, \xi) &= \eta^T x - \xi = (\eta^T, \xi) \begin{pmatrix} x \\ -1 \end{pmatrix} \\ &= \tilde{\zeta}^T (D^T x - d) + \mu^T x - \mu_{n+1}. \end{aligned} \quad (2.54)$$

It follows that  $\zeta(x, \eta, \xi)$  is normally distributed with

$$\begin{aligned}
\mathbb{E}[\zeta(x, \eta, \xi)] &= \mu^T x - \mu_{n+1} \\
\mathbb{V}\text{ar}[\zeta(x, \eta, \xi)] &= \|D^T x - d\|^2 \\
&= x^T D D^T x - 2(Dd)^T x + \|d\|^2,
\end{aligned} \tag{2.55}$$

where  $\|\cdot\|$  denotes the Euclidean norm. The first term on the right-hand-side is the variance of  $\eta^T x$  with  $DD^T$  being the covariance matrix of  $\eta$ . In the second term  $(Dd)^T x$  is the covariance between  $\eta^T x$  and  $\xi$  with  $Dd$  being the cross-covariance vector between  $\eta$  and  $\xi$ . The third term is the variance of  $\xi$ .

If  $\mathbb{V}\text{ar}[\zeta(x, \eta, \xi)] = 0$  then  $\zeta(x, \eta, \xi) = \mathbb{E}[\zeta(x, \eta, \xi)]$ , a.s., otherwise the standardized  $\zeta(x, \eta, \xi)$  has a standard normal distribution.

In the case  $\|D^T x - d\| > 0$  we obtain via standardization

$$\begin{aligned}
G(x) &= \mathbb{P}(\zeta(x, \eta, \xi) \geq 0) = 1 - \mathbb{P}(\zeta(x, \eta, \xi) \leq 0) \\
&= 1 - \mathbb{P}\left(\frac{\zeta(x, \eta, \xi) - \mathbb{E}[\zeta(x, \eta, \xi)]}{\|D^T x - d\|} \leq \frac{-\mu^T x + \mu_{n+1}}{\|D^T x - d\|}\right) \\
&= 1 - \Phi\left(\frac{-\mu^T x + \mu_{n+1}}{\|D^T x - d\|}\right) = \Phi\left(\frac{\mu^T x - \mu_{n+1}}{\|D^T x - d\|}\right),
\end{aligned} \tag{2.56}$$

where in the last step we utilized the symmetry of the standard normal distribution, that means, we made use of the relation  $\Phi(x) = 1 - \Phi(-x)$ ,  $\forall x \in \mathbb{R}$ . Thus we get the following formula for  $G(x)$ :

$$G(x) = \begin{cases} 1, & \text{if } D^T x - d = 0 \\ & \text{and } \mu^T x - \mu_{n+1} \geq 0, \\ 0, & \text{if } D^T x - d = 0 \\ & \text{and } \mu^T x - \mu_{n+1} < 0, \\ \Phi\left(\frac{\mu^T x - \mu_{n+1}}{\|D^T x - d\|}\right), & \text{if } D^T x - d \neq 0. \end{cases} \tag{2.57}$$

Regarding the constraint  $G(x) \geq \alpha$ , under the assumption  $D^T x - d \neq 0$  we get

$$\begin{aligned}
G(x) \geq \alpha &\iff \Phi\left(\frac{\mu^T x - \mu_{n+1}}{\|D^T x - d\|}\right) \geq \alpha \\
&\iff \Phi^{-1}(\alpha) \|D^T x - d\| - \mu^T x \leq -\mu_{n+1}.
\end{aligned} \tag{2.58}$$

In the case when  $D^T x - d = 0$  holds, the last inequality reduces to the first case in (2.57), consequently the equivalence holds in all cases. Note that for  $\alpha \geq \frac{1}{2}$  we have  $\Phi^{-1}(\alpha) \geq 0$ . The Euclidean norm being convex,  $\|D^T x - d\|$  is a convex function of  $x$ . Consequently, assuming that  $\alpha \geq \frac{1}{2}$  holds, the function on the left-hand-side of the last inequality in (2.58) is a convex function. This implies that the set of feasible solutions w.r. to this constraint is a convex set. We have derived the following theorem:

**Theorem 2.5.** *Kataoka 1963 [178], Van de Panne and Popp 1963 [333]. Let the  $(n+1)$ -dimensional random vector  $\zeta^T = (\eta^T, \xi)^T$  have a multivariate normal distribution and let  $\alpha \geq \frac{1}{2}$ . Then the set  $\mathcal{B}(\alpha) = \{x \mid G(x) \geq \alpha\}$  is convex.*

For the case, when  $\alpha < \frac{1}{2}$  holds, we have the following assertion:

**Theorem 2.6.** *Kall 1976 [154]. Let  $n > 1$  and assume that the  $(n+1)$ -dimensional random vector  $\zeta^T = (\eta^T, \xi)^T$  has a non-degenerate multivariate normal distribution. If  $\alpha < \frac{1}{2}$  then either  $\mathcal{B}(\alpha) = \mathbb{R}^n$  holds or otherwise  $\mathcal{B}(\alpha)$  is a non-convex set.*

*Proof:* Let  $\hat{x} \in \mathbb{R}^n$  be such that  $\hat{x} \notin \mathcal{B}(\alpha)$  holds. We will show, that under our assumptions, there exist  $x^{(1)} \in \mathcal{B}(\alpha)$  and  $x^{(2)} \in \mathcal{B}(\alpha)$  such that  $x^{(1)} \neq x^{(2)}$  and  $\hat{x} = \frac{1}{2}(x^{(1)} + x^{(2)})$  holds. From this our assertion follows immediately.

$n > 1$  implies that there exists  $v \in \mathbb{R}^n$  such that  $v \neq 0$  and  $\mu^T v = 0$  hold. Let us consider the constraint (2.58) along the line  $x(\lambda) = \hat{x} + \lambda v$ ,  $\lambda \in \mathbb{R}$ :

$$\Phi^{-1}(\alpha) \|D^T x(\lambda) - d\| - \mu^T \hat{x} \leq -\mu_{n+1},$$

where we used that  $\mu^T x(\lambda) = \mu^T \hat{x}$ ,  $\forall \lambda \in \mathbb{R}$  holds. We obviously have  $\|D^T x(\lambda) - d\| \geq \|D^T x(\lambda)\| - \|d\|$ , and an easy computation yields

$$\|D^T x(\lambda)\|^2 = \lambda^2 v^T D D^T v + 2\lambda v^T D D^T \hat{x} + \hat{x}^T D D^T \hat{x}.$$

Matrix  $D$  has full row rank and  $v \neq 0$ , therefore  $\lim_{\lambda \rightarrow \pm\infty} \|D^T x(\lambda) - d\| = \infty$  holds.

Taking into account  $\Phi^{-1}(\alpha) < 0$ , this implies that  $\exists \lambda_0 \in \mathbb{R}$ , such that both  $x(\lambda_0) \in \mathcal{B}(\alpha)$  and  $x(-\lambda_0) \in \mathcal{B}(\alpha)$ . Obviously  $x(\lambda_0) \neq x(-\lambda_0)$  and  $\hat{x} = \frac{1}{2}(x(\lambda_0) + x(-\lambda_0))$ .  $\square$

For the probability function with reversed random inequalities, that means, for  $\hat{G}(x) := \mathbb{P}(\eta^T x \leq \xi) = \mathbb{P}(\zeta(x, \eta, \xi) \leq 0)$  we get

$$\hat{G}(x) = \begin{cases} 1, & \text{if } D^T x - d = 0 \text{ and } \mu^T x - \mu_{n+1} \leq 0 \\ 0, & \text{if } D^T x - d = 0 \text{ and } \mu^T x - \mu_{n+1} > 0 \\ \Phi\left(\frac{-\mu^T x + \mu_{n+1}}{\|D^T x - d\|}\right), & \text{if } D^T x - d \neq 0. \end{cases} \quad (2.59)$$

This can either be derived by an analogous argumentation as above, or more directly as follows. Observe that if  $\zeta(x, \eta, \xi)$  has a normal distribution, then  $-\zeta(x, \eta, \xi)$  also has a normal distribution with the same variance and with reversed sign of the expected value. Thus (2.57) can be directly applied for  $-\zeta(x, \eta, \xi)$ , by writing  $\hat{G}$  as  $\hat{G}(x) = \mathbb{P}(-\zeta(x, \eta, \xi) \geq 0)$ .

Utilizing the formulas (2.57) and (2.59), we obtain the following equivalent representations of probability constraints:

$$\begin{aligned} \mathbb{P}(\eta^T x \geq \xi) \geq \alpha &\iff \Phi^{-1}(\alpha) \|D^T x - d\| - \mu^T x \leq -\mu_{n+1} \\ \mathbb{P}(\eta^T x \leq \xi) \geq \alpha &\iff \Phi^{-1}(\alpha) \|D^T x - d\| + \mu^T x \leq \mu_{n+1} \end{aligned} \quad (2.60)$$



where for  $\alpha \geq \frac{1}{2}$  the functions on the left-hand-side of the equivalent inequalities are convex, therefore the feasible domain determined by these inequalities is convex.

We turn our attention to the case with reverse inequalities in the constraints, that means, we deal with  $G(x) \leq \beta$  and  $\widehat{G}(x) \leq \beta$ . In the case when the probability distribution is degenerate, the previously used technique for deriving the equivalent form leads to strict inequalities. Having  $D^T x - d = 0$ , the formulas (2.57) and (2.59) imply a strict inequality (the second cases in these formulas apply). Assuming non-degeneracy of the probability distribution, we obtain the following equivalent representations by reversing the inequalities in (2.60):

$$\begin{aligned} \mathbb{P}(\eta^T x \geq \xi) &\leq \beta \iff \Phi^{-1}(\beta) \|D^T x - d\| - \mu^T x \geq -\mu_{n+1} \\ \mathbb{P}(\eta^T x \leq \xi) &\leq \beta \iff \Phi^{-1}(\beta) \|D^T x - d\| + \mu^T x \geq \mu_{n+1} \end{aligned} \quad (2.61)$$

where, provided that  $\beta \leq \frac{1}{2}$  holds, the functions on the left-hand-side of the equivalent inequalities are concave, consequently the feasible domain determined by these inequalities is convex.

In the case when the probability distribution is degenerate, we observe a similar asymmetry as in the previous section on page 98 between the two formulations differing in the direction of the inequality ( $G(x) \geq \alpha$  versus  $G(x) \leq \beta$ ). The remedy is analogous: In practical modeling this difficulty can usually be overcome by working with strict inequalities in the model formulation. For instance, taking the constraint  $\mathbb{P}_\xi(\zeta(x, \xi) < 0) \leq \beta$ , this can be equivalently formulated as

$$\mathbb{P}_\xi(\zeta(x, \xi) \geq 0) \geq 1 - \beta$$

which results according to (2.60) in the linear constraint

$$\Phi^{-1}(1 - \beta) \|D^T x - d\| - \mu^T x \leq -\mu_{n+1}$$

thus determining a convex feasible domain for  $\beta \leq \frac{1}{2}$ .

Next we turn our attention to models with probability functions in the objective and restrict our discussion to the case, when  $\zeta^T = (\eta^T, \xi)^T$  has a *non-degenerate* multivariate normal distribution. The distribution of  $\zeta$  is non-degenerate, if and only if the matrix  $\begin{pmatrix} D \\ d^T \end{pmatrix}$  has full row rank, see Definition 2.14. and (2.53). Consequently, in the non-degenerate case  $D^T x - d \neq 0$  holds for all  $x \in \mathbb{R}^n$ . In particular, choosing  $x = 0$  shows that  $d \neq 0$  holds.

In the non-degenerate case we have, see (2.57):

$$G(x) = \Phi\left(\frac{\mu^T x - \mu_{n+1}}{\|D^T x - d\|}\right) \quad \forall x \in \mathbb{R}^n. \quad (2.62)$$

In a maximization problem the desired property of  $G(x)$  would be pseudo-concavity. Unfortunately,  $G(x)$  is not even quasi-concave. Quasi-concavity is namely equivalent with the convexity of all of the upper level sets (see page 83). This is implied by (2.60) for  $\alpha \geq \frac{1}{2}$ . For any  $0 < \alpha < \frac{1}{2}$ , however, the lower level set is convex accord-

ing to (2.61). The upper level sets corresponding to the same  $\alpha$  cannot be also convex, because this would mean that both the upper- and the lower level sets are half-spaces. This is not possible due to our non-degeneracy assumption  $\|D^T x - d\| \neq 0$  for all  $x \in \mathbb{R}^n$ . Consequently  $G(x)$  is not quasi-concave. An analogous reasoning shows that  $G(x)$  is not quasi-convex, either.

Introducing the notation

$$g(x) = \frac{\mu^T x - \mu_{n+1}}{\|D^T x - d\|}$$

we get  $G(x) = \Phi(g(x))$ . Fortunately, by restricting  $G(x)$  to certain half-spaces we have

**Proposition 2.36.** *If  $\zeta^T = (\eta^T, \xi)^T$  has a non-degenerate multivariate normal distribution, then both  $g(x)$  and  $G(x)$  are*

- a) pseudo-concave on the half-space  $\{x \mid \mu^T x \geq \mu_{n+1}\}$  and
- b) pseudo-convex on the half-space  $\{x \mid \mu^T x \leq \mu_{n+1}\}$ .

*Proof:* Due to the non-degeneracy assumption  $\|D^T x - d\| > 0, \forall x \in \mathbb{R}^n$  holds. Due to Proposition 2.32. on page 83, the fractional function  $g(x)$  is pseudo-concave on convex sets where the numerator is nonnegative, and pseudo-convex on convex sets where the numerator is non-positive. From this the result regarding  $g(x)$  follows. Utilizing the fact that  $\Phi$  is a strictly monotonically increasing, differentiable function, with  $\Phi'(x) \neq 0 \forall x \in \mathbb{R}$ , the assertion concerning  $G(x)$  follows from the already proved assertion regarding  $g(x)$  and from Proposition 2.33. on page 84.  $\square$

Let us consider (2.36) on page 90, which in our case has the form

$$\begin{aligned} & \max \quad \Phi \left( \frac{\mu^T x - \mu_{n+1}}{\|D^T x - d\|} \right) \\ & \text{s.t.} \quad x \in \mathcal{B}. \end{aligned} \tag{2.63}$$

According to Proposition 2.36., the objective function of this linearly constrained problem is pseudo-concave, if  $x \in \mathcal{B}$  implies  $\mu^T x \geq \mu_{n+1}$ . Thus, in this case, (2.63) is a convex programming problem. Taking into account the strict monotonicity of  $\Phi$ , (2.63) is equivalent to the following linearly constrained convex programming problem

$$\begin{aligned} & \max \quad \frac{\mu^T x - \mu_{n+1}}{\|D^T x - d\|} \\ & \text{s.t.} \quad x \in \mathcal{B}. \end{aligned} \tag{2.64}$$

This problem belongs to the class of fractional programming problems, see, for instance, Avriel, Diewert, Schaible, and Zang [8] and Schaible [297]. Proposition 2.36. implies that the objective function in (2.64) is pseudo-concave in the half-space  $\{x \mid \mu^T x \geq \mu_{n+1}\}$  and it is pseudo-convex in the half-space  $\{x \mid \mu^T x \leq \mu_{n+1}\}$ .

Consequently, if  $\mu^T x \geq \mu_{n+1} \forall x \in \mathcal{B}$  holds, then (2.64) is a convex programming problem. This property can be enforced, for instance, by including a linear inequality of the form  $\mu^T x \geq \mu_{n+1}$  into the definition of  $\mathcal{B}$ . This might be well justified if a high probability is to be achieved by maximizing  $\mathbb{P}(\eta^T x \geq \xi)$ . For achieving high probabilities it is necessary to have  $\mathbb{E}[\eta^T x] \geq \mathbb{E}[\xi]$ , which is just the required inequality.

If the reverse inequality  $\{x \mid \mu^T x \leq \mu_{n+1}\}$  holds over  $\mathcal{B}$ , then our objective is pseudo-convex, (2.64) involves maximizing a pseudo-convex function, and thus it becomes much more difficult to solve numerically. In the general case, when none of the two inequalities involving expectations holds uniformly over  $\mathcal{B}$ , then (2.64) becomes a general non-convex optimization problem. In this case efficient solution methods are only available for rather low dimensions of  $x$ .

In the case when (2.63) and (2.64) are formulated as minimization problems, the above results can be adapted in a straightforward manner. If we take  $\hat{G}(x) = \mathbb{P}(\eta^T x \leq \xi)$  instead of  $G(x)$  then the above discussion applies with exchanged roles of the inequalities  $\mu^T x \geq \mu_{n+1}$  and  $\mu^T x \leq \mu_{n+1}$ .

Finally we discuss the special case when  $\xi$  is deterministic. Note that the non-degeneracy assumption above implies that all components of  $\eta$  as well as  $\xi$  have non-degenerate univariate marginal distributions, that means, both the “technology matrix” and the right-hand-side are stochastic. We assume now that  $\xi \equiv \mu_{n+1} := h$  holds with  $h \in \mathbb{R}$  being deterministic. Considering (2.54), this means that  $d = 0$  holds throughout. Non-degeneracy of the distribution in this case means that  $D$  has full row rank.

The explicit form of  $\hat{G}$  and the probability constraint can simply be obtained by setting  $d = 0$  in (2.59) and in (2.61), respectively. Considering the problem of minimizing  $\hat{G}(x)$  results in:

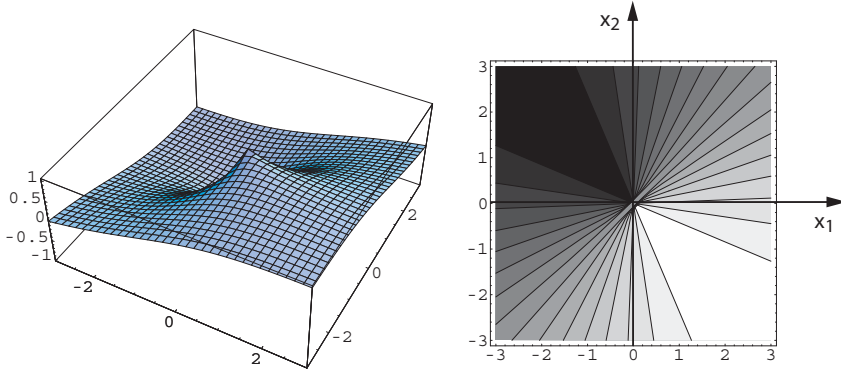
$$\begin{aligned} \min \quad & \frac{-\mu^T x + h}{\|D^T x\|} \\ \text{s.t.} \quad & x \in \mathcal{B} \end{aligned} \quad (2.65)$$

which makes only sense under the assumption  $0 \notin \mathcal{B}$ . We have seen that problem (2.65) is a convex programming problem provided that  $\mu^T x \geq h, \forall x \in \mathcal{B}$  holds.

Figure 2.6 shows the graph and the contour lines of the function

$$f(x_1, x_2) = \frac{x_1 - x_2}{\sqrt{(x_1 + x_2)^2 + (x_1 - x_2)^2}}$$

which is the quotient of a linear and a convex function. In the contour plot darker regions represent lower values. Let  $\varepsilon > 0$ ; for the figure we have chosen  $\varepsilon = 0.1$ . The function  $f$  is pseudo-concave for  $\{x \in \mathbb{R}^2 \mid x_1 \geq x_2 + \varepsilon\}$  and pseudo-convex for  $\{x \in \mathbb{R}^2 \mid x_1 \leq x_2 - \varepsilon\}$ .



**Fig. 2.6** Quotient of a linear and a nonlinear convex function.

### Stable distributions

In the previous section, in the derivation of the explicit formula (2.58), it seems to be essential at a first glance, that both the expected value and the variance exist for  $\zeta(x, \eta, \xi)$ . A more careful analysis reveals, however, that quite other properties of the normal distribution are those, which matter.

Before carrying out this analysis, we discuss classifications of univariate distributions, which will be needed later on. We define a relation  $\diamond$  between univariate distribution functions, see Feller [92]. Let  $F$  and  $H$  be two univariate distribution functions, then

$$F \diamond H \iff \exists a > 0, b : H(x) = F(ax + b) \quad \forall x \in \mathbb{R}^1 \text{ holds} \quad (2.66)$$

or equivalently

$$F \diamond H \iff \exists a > 0, b : H\left(\frac{x-b}{a}\right) = F(x) \quad \forall x \in \mathbb{R}^1. \quad (2.67)$$

This relation is obviously reflexive, symmetric, and transitive. Consequently we obtain a classification of all distribution functions. We may choose a representative from each class, and consider it as a *standard distribution* for that class. Let  $\mathcal{D}$  be a class in this classification, and let  $H_0$  be the standard distribution in  $\mathcal{D}$ . Then for any  $F \in \mathcal{D}$  we have:  $\exists a > 0, b$ , such that  $F(x) = H_0(\frac{x-b}{a})$ ,  $\forall x \in \mathbb{R}^1$  holds.  $a$  is called the *scale*- and  $b$  the *location* parameter of  $F$  (w.r. to the standard distribution). The classes in this classification are also called *location-scale classes*.

Let  $\zeta$  be a random variable with  $F_\zeta \in \mathcal{D}$ . This fact will be denoted as  $\zeta \sim \mathcal{D}$ . Then  $\exists a, b \in \mathbb{R}, a > 0$  such that  $F_\zeta(x) = H_0(\frac{x-b}{a})$  holds. This relation has the following interpretation: Let  $\chi = \frac{\zeta-b}{a}$ . Then we have

$$\mathbb{P}(\chi \leq x) = \mathbb{P}(\zeta \leq ax + b) = F_\zeta(ax + b) = H_0(x),$$

that means,  $\chi$  has the standard distribution of  $\mathcal{D}$ . The transformation above is called *standardization* of  $\zeta$ . This can also be expressed as follows: for any  $\zeta \sim \mathcal{D} \exists a > 0, b$ , such that  $\zeta = a\chi + b$  and  $F_\chi = H_0$  holds. A final remark to this concept: let  $\zeta$  be a random variable with  $F_\zeta \in \mathcal{D}$ , and let  $p > 0, q$  be real numbers. Then obviously  $p\zeta + q \sim \mathcal{D}$  holds.

We consider next the set of symmetric distributions. A distribution is called *symmetric* if for the distribution function  $F$  the following relation holds (see Feller [92]):  $F(x) = 1 - F_-(-x) \forall x \in \mathbb{R}$ , where  $F_-(-x)$  stands for the left-sided limit of  $F$  at  $-x$ . If the density function  $f$  exists then the condition for symmetry can be written as  $f(x) = f(-x) \forall x \in \mathbb{R}$ . On the set of symmetric distributions the following equivalence relation establishes a classification:

$$F \diamond H \iff \exists a > 0 : H(x) = F(ax) \forall x \in \mathbb{R}^1.$$

The classes in this classification will be called *symmetric scale classes*.

If the random variable  $\zeta$  has a symmetric distribution, this is clearly equivalent with  $\zeta$  and  $-\zeta$  having the same distribution function, that means, with  $F_\zeta = F_{-\zeta}$ . Let  $\mathcal{S}$  be a class of symmetric distributions. Then  $\zeta \sim \mathcal{S}$  implies  $p\zeta \sim \mathcal{S} \forall p \in \mathbb{R}, p \neq 0$ . For  $p > 0$  this is clear from the definition. If  $p < 0$  then we may write  $p\zeta = (-p)(-\zeta)$ . Now we have  $F_\zeta = F_{-\zeta}$ , and the assertion follows immediately.

If a location-scale class  $\mathcal{D}$  contains a single symmetric distribution, then it obviously contains the whole symmetric scale class  $\mathcal{S}$  of this distribution. In this case the standard distribution can be selected as a symmetric distribution, that means,  $H_0 \in \mathcal{S}$ . Let  $\zeta \sim \mathcal{D}$  and  $p, q \in \mathbb{R}, p \neq 0$ . Then, for such classes,  $p\zeta + q \sim \mathcal{D}$  holds. For  $p > 0$  this is clear from the definition. Let us assume that  $p < 0$  holds. Standardization gives that  $\exists a > 0, b$  such that  $\zeta = a\chi + b$  and  $F_\chi = H_0$ . Substitution results in  $p\zeta + q = ap\chi + bp + q$ . From this follows  $ap\chi \sim \mathcal{S}$  and consequently  $p\zeta + q \sim \mathcal{D}$ .

Let us introduce the notion of a stable distribution next. For this concept see, for instance, Feller [92] and Uchaikin and Zolotarev [330].

A distribution function  $F$ , the corresponding probability distribution, and a random variable having this distribution are called *stable*, if for any real numbers  $s_1 > 0, m_1, s_2 > 0$ , and  $m_2$  there exist real numbers  $s > 0$  and  $m$ , such that

$$F\left(\frac{x - m_1}{s_1}\right) * F\left(\frac{x - m_2}{s_2}\right) = F\left(\frac{x - m}{s}\right), \quad \forall x \in \mathbb{R}^1 \quad (2.68)$$

holds, where  $*$  stands for the convolution operator. Let  $F$  be a stable distribution function and let  $\mathcal{D}$  be its class in the above classification. From (2.68) immediately follows, that all  $H \in \mathcal{D}$  are stable, that means, we may use the term *class of stable distributions*. In particular, the standard distribution  $H_0 \in \mathcal{D}$  is also stable. Another easy consequence of (2.68) is the following: if  $F \in \mathcal{D}, H \in \mathcal{D}$ , and  $\mathcal{D}$  is a stable class, then  $F * H \in \mathcal{D}$  holds. Using the fact, that the distribution function of the sum of two stochastically independent random variables is the convolution of their distribution functions, we get the following: Let  $\mathcal{D}$  be a stable class,  $\zeta_i \sim \mathcal{D}, i = 1, \dots, s, \lambda_i \in \mathbb{R} \lambda_i > 0 \forall i$ . Assume that  $\zeta_i, i = 1, \dots, s$  are stochastically independent. Then

the distribution function of  $\sum_{i=1}^s \lambda_i \zeta_i$  also belongs to  $\mathcal{D}$ . This property is, however, not sufficient for our purposes: in (2.53) we deal with arbitrary linear combinations of independent random variables.

A distribution function  $F$ , the corresponding probability distribution, and a random variable having this distribution are called *strictly stable*, if for any real numbers  $s_1 > 0$  and  $s_2 > 0$  there exists a real number  $s > 0$ , such that

$$F\left(\frac{x}{s_1}\right) * F\left(\frac{x}{s_2}\right) = F\left(\frac{x}{s}\right), \quad \forall x \in \mathbb{R}^1 \quad (2.69)$$

holds, where  $*$  stands as before for the convolution operator. In the following we restrict our attention to symmetric distributions. Let  $F$  be a strictly stable distribution function and let  $\mathcal{S}$  be its class in the classification of symmetric distributions. The analogous results hold, as for stable distributions. In particular, if  $F \in \mathcal{S}$  and  $H \in \mathcal{S}$ , then  $F * H \in \mathcal{S}$  follows. This implies for symmetric distributions the following: Let  $\mathcal{S}$  be a strictly stable class of symmetric distributions,  $\zeta_i \in \mathcal{S}$ ,  $i = 1, \dots, s$ ,  $\lambda_i \in \mathbb{R} \forall i$ , and not all  $\lambda_i$ 's are zero. Assume that  $\zeta_i$ ,  $i = 1, \dots, s$  are stochastically independent. Then the distribution function of  $\sum_{i=1}^s \lambda_i \zeta_i$  also belongs to  $\mathcal{S}$ .

As an example for a stable class of distributions let us shortly discuss the univariate normal distribution. The univariate normal distribution functions form a location-scale class, because they are of the form:  $F(x) = \Phi\left(\frac{x-b}{a}\right)$ ,  $0 < a \in \mathbb{R}$ ,  $b \in \mathbb{R}$ , where  $\Phi$  is the distribution function of the standard normal distribution. This is a stable class. To see this, it is sufficient to check the stability of  $\Phi$ . Considering the convolution (2.68)

$$\Phi\left(\frac{x-m_1}{\sigma_1}\right) * \Phi\left(\frac{x-m_2}{\sigma_2}\right) = \Phi\left(\frac{x-m}{s}\right), \quad \forall x \in \mathbb{R}^1,$$

where the left-hand-side is the distribution function of the sum of two independent  $\xi \sim \mathcal{N}(m_1, \sigma_1^2)$  and  $\eta \sim \mathcal{N}(m_2, \sigma_2^2)$  random variables. We know that  $\xi + \eta$  has a normal distribution. On the other hand, the expected value is additive w.r. to summation, and the variance is also additive provided that the random variables are stochastically independent. Therefore the above relation holds for  $m = m_1 + m_2$  and  $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$ . This argumentation also shows that the class of symmetric (centered) normal distribution functions  $F(x) = \Phi\left(\frac{x}{a}\right)$ ,  $0 < a \in \mathbb{R}$  form a strictly stable class of symmetric distributions.

Now we take the proposed second look at the derivation of the explicit form for  $G$  in Section 2.2.3.

1. The multivariate distribution of  $\zeta$  was defined by the affine linear relations (2.53) for the realizations, in terms of the *i.i.d.* (independent and identically distributed) random variables  $\tilde{\zeta}_i$ ,  $i = 1 \dots s$ . In that particular case the distribution of  $\tilde{\zeta}_i$  was

standard normal,  $\forall i$ , which, as discussed above, belongs to the strictly stable class of symmetric normal distributions.

2. Subsequently we have established in (2.54) an affine linear relation for  $\zeta(x, \eta, \xi)$ , in terms of  $\tilde{\xi}$ .
3. Considering the linear part, this is a linear combination of random variables with distributions from a strictly stable class, therefore the linear combination belongs also to that class. Due to the additive deterministic term,  $\zeta(x, \eta, \xi)$  belongs to the stable class of normal distributions. In addition, using the specific properties of the normal distribution, we were also able to compute the parameters of  $\zeta(x, \eta, \xi)$ , in terms of our decision variables  $x$ .
4. Finally, in (2.54), we have standardized  $\zeta(x, \eta, \xi)$  in order to derive a formula for  $G(x)$ , involving the distribution function of the standard distribution in the location–scale class. Using this formula, the constraint  $G(x) \geq \alpha$  has been reformulated as (2.58). By good luck, this resulted in a constraint of the convex programming type.

Another well-known stable univariate distribution is the Cauchy distribution, see, for instance, Feller [92]. For this distribution the expected value and consequently the variance do not exist. The density function of the Cauchy distribution  $\mathcal{C}(m, t)$  is the following:

$$f(x) = \frac{1}{\pi} \frac{t}{t^2 + (x - m)^2}, \quad -\infty < x < \infty,$$

where  $m$  is a location parameter and  $t > 0$  is a scale parameter. Taking  $m = 0$  the resulting subclass of symmetric distributions is strictly stable. The distribution function of the standard Cauchy distribution  $\mathcal{C}(0, 1)$ , defined by the density function with  $t = 1$

$$\psi(x) = \frac{1}{\pi} \frac{1}{1 + x^2}, \quad -\infty < x < \infty,$$

will be denoted by  $\Psi$ . The following fact is also well-known, see, for instance, Feller [92]: Let  $\xi \sim \mathcal{C}(m_1, t_1)$  and  $\eta \sim \mathcal{C}(m_2, t_2)$ , and assume that  $\xi$  and  $\eta$  are stochastically independent. Then  $\xi + \eta \sim \mathcal{C}(m_1 + m_2, t_1 + t_2)$  holds.

We will carry out the above procedure for a multivariate Cauchy distribution, see Marti [222].

**Definition 2.15.** *The  $r$ -dimensional random vector  $\zeta$  has a non-degenerate multivariate Cauchy distribution, if there exist an  $(r \times s)$  matrix  $B$  with full row rank and having at least one nonzero in each of its columns and  $m \in \mathbb{R}^r$ , such that*

$$\zeta = B\tilde{\zeta} + m, \tag{2.70}$$

where  $\tilde{\zeta}$  is an  $s$ -dimensional random vector with its components being stochastically independent and  $\tilde{\zeta}_i$  having a standard Cauchy distribution,  $\forall i$ .

Let us assume, that the  $(n + 1)$ -dimensional random vector  $\zeta^T = (\eta^T, \xi)^T$  has a non-degenerate multivariate Cauchy distribution. In the same way, as in Section 2.2.3, we get:

$$\zeta(x, \eta, \xi) = \tilde{\zeta}^T(D^T x - d) + m^T x - m_{n+1}. \quad (2.71)$$

Let us remark that  $\|D^T x - d\| \neq 0$  holds for all  $x$ , due to the assumption that the transformation matrix  $\begin{pmatrix} D \\ d^T \end{pmatrix}$  has full row rank (see also (2.53)). We conclude that  $\zeta(x, \eta, \xi)$  has a Cauchy distribution, and proceed by computing its parameters. If  $(D^T x - d)_i \neq 0$  then  $\tilde{\zeta}_i(D^T x - d)_i \sim \mathcal{C}(0, |(D^T x - d)_i|)$  holds. Consequently we have  $\tilde{\zeta}^T(D^T x - d) \sim \mathcal{C}(0, \|D^T x - d\|_1)$ , where for  $y \in \mathbb{R}^s$   $\|y\|_1 := \sum_{i=1}^s |y_i|$ . Finally we get:

$$\zeta(x, \eta, \xi) \sim \mathcal{C}(m^T x - m_{n+1}, \|D^T x - d\|_1).$$

Using standardization, as in (2.57) we get the following formula for  $G(x)$ :

$$G(x) = 1 - \Psi\left(\frac{-m^T x + m_{n+1}}{\|D^T x - d\|_1}\right) = \Psi\left(\frac{m^T x - m_{n+1}}{\|D^T x - d\|_1}\right), \quad (2.72)$$

where we utilized the symmetry of the standard Cauchy distribution. Comparing this with the analogous formula (2.57) for the non-degenerate multivariate normal distribution, it can be observed that the sole difference is the different norm in the denominator.

We proceed now analogously as in (2.58) to arrive at:

$$G(x) \geq \alpha \iff \Psi^{-1}(\alpha) \|D^T x - d\|_1 - m^T x \leq -m_{n+1}. \quad (2.73)$$

The standard Cauchy distribution being symmetric, for  $\alpha \geq \frac{1}{2}$ ,  $\Psi^{-1}(\alpha) \geq 0$  holds. Because norms are convex functions,  $\|D^T x - d\|_1$  is a convex function of  $x$ . As for the normal distribution, we conclude that the function on the left-hand-side of the inequality is a concave function, and the set of  $x$  vectors, for which this inequality holds, is convex. We have derived the following theorem:

**Theorem 2.7.** *Marti 1971 [222]. Let the  $(n+1)$ -dimensional random vector  $\zeta^T = (\eta^T, \xi)^T$  have a non-degenerate multivariate Cauchy distribution and let  $\alpha \geq \frac{1}{2}$ . Then the set  $\mathcal{B}(\alpha) = \{x \mid G(x) \geq \alpha\}$  is convex.*

The alternative formulations of the probability constraints are analogous to those for the normal distribution. The difference is that, instead of the Euclidean norm, the  $\|\cdot\|_1$ -norm is to be substituted throughout. This seems to introduce, however, an additional difficulty: the  $\|\cdot\|_1$ -norm is a non-differentiable function of its argument. Under the assumption  $\alpha \geq \frac{1}{2}$ , a second look reveals, however, that by introducing additional variables the constraint (2.73) can be equivalently formulated as a set of linear constraints. In this respect, probability constraints are easier to deal with for the Cauchy distribution as for the normal distribution. For discussing the transformation let us formulate (2.73) in a detailed form:

$$\Psi^{-1}(\alpha) \sum_{i=1}^s |D_i^T x - d_i| - m^T x \leq -m_{n+1}, \quad (2.74)$$



where  $D_i$  is the  $i^{\text{th}}$  column of  $D$ . This constraint is equivalent to the following system of linear constraints:

$$\begin{aligned} -m^T x + \Psi^{-1}(\alpha) \sum_{i=1}^s y_i &\leq -m_{n+1} \\ D_i^T x - y_k &\leq d_k, \quad k = 1, \dots, s \\ D_i^T x + y_k &\geq d_k, \quad k = 1, \dots, s, \end{aligned} \quad (2.75)$$

in the following sense: Let  $\bar{x}$  be a feasible solution of (2.74). Choosing  $\bar{y}_k = |D_k^T \bar{x} - d_k| \forall k$  implies that  $(\bar{x}, \bar{y}_k, k = 1, \dots, s)$  is a feasible solution of (2.74). Vice versa, let  $(\hat{x}, \hat{y}_k, k = 1, \dots, s)$  be feasible for (2.75). Then the inequality  $|D_k^T \hat{x} - d_k| \leq y_k$  holds  $\forall k$ , which implies that  $\hat{x}$  is feasible for (2.74).

There is an important special case, as observed by Marti [222], in which the problem transforms into a deterministic LP problem, without introducing additional variables and constraints. Let us assume that  $\mathcal{B} \subset \mathbb{R}_+^n$  holds which is the case, for instance, if the system of linear inequalities defining  $\mathcal{B}$  includes  $x \geq 0$ . Assume further, that the components of  $(\eta, \xi)$  are stochastically independent and that they have Cauchy distributions  $\eta_i \sim \mathcal{C}(m_i, t_i)$   $i = 1, \dots, n$  and  $\xi \sim \mathcal{C}(m_{n+1}, t_{n+1})$ . In this case the matrix  $\begin{pmatrix} D \\ d^T \end{pmatrix}$  is a diagonal  $((n+1) \times (n+1))$  matrix, with the  $t_i$ 's on its diagonal, see (2.2.3). Consequently we get  $\|D^T x - d\|_1 = \sum_{i=1}^n t_i x_i + t_{n+1}$  and (2.73) becomes a linear constraint.

### A distribution-free approach

The sole assumption in this section is that the second moments of  $(\eta^T, \xi)$  exist. Let  $(\mu^T, \mu_{n+1}) = \mathbb{E}[(\eta^T, \xi)]$  and  $\Sigma$  be the covariance matrix of  $(\eta^T, \xi)$ . We assume that  $\Sigma$  is positive definite and take the Cholesky factorization  $\Sigma = LL^T$  with  $L$  being a lower triangular matrix (cf. the discussion on page 99). We consider  $L$  in the partitioned form

$$L = \begin{pmatrix} D \\ d^T \end{pmatrix},$$

where  $D$  is an  $(n \times n)$  matrix and  $d \in \mathbb{R}^n$ . For  $\zeta(x, \eta, \xi) = \eta^T x - \xi$  we get the same expression (2.55) as for the normal distribution

$$\begin{aligned} \mathbb{E}[\zeta(x, \eta, \xi)] &= \mu^T x - \mu_{n+1} \\ \mathbb{V}\text{ar}[\zeta(x, \eta, \xi)] &= \|D^T x - d\|^2. \end{aligned}$$

The general idea is to employ upper bounds on the probability function  $G(x) = \mathbb{P}(\eta^T x - \xi \geq 0)$ . Utilizing the Chebyshev-inequality we get

$$\begin{aligned}
G(x) &= \mathbb{P}((\eta - \mu)^T x - (\xi - \mu_{n+1}) \geq -\mu^T x + \mu_{n+1}) \\
&\leq \mathbb{P}(|(\eta - \mu)^T x - (\xi - \mu_{n+1})| \geq -\mu^T x + \mu_{n+1}) \\
&\leq \frac{\text{Var}(\eta^T x - \xi)}{(-\mu^T x + \mu_{n+1})^2} = \frac{\|D^T x - d\|^2}{(-\mu^T x + \mu_{n+1})^2}.
\end{aligned} \tag{2.76}$$

We consider the probability constraint  $G(x) \leq \beta$  with  $\beta$  small, for instance,  $\beta = 0.01$ . The idea is to require instead of this inequality the stronger inequality

$$\frac{\|D^T x - d\|^2}{(-\mu^T x + \mu_{n+1})^2} \leq \beta. \tag{2.77}$$

For having a nonempty solution set of this inequality, for small  $\beta$  values we may suppose that  $-\mu^T x + \mu_{n+1} > 0$  holds. This may be enforced by including a constraint  $-\mu^T x + \mu_{n+1} > \varepsilon$ , with  $\varepsilon > 0$ , into the set of linear constraints of the problem. Assuming this, we can write (2.77) as follows

$$\beta^{-\frac{1}{2}} \|D^T x - d\| + \mu^T x \leq \mu_{n+1} \tag{2.78}$$

which defines a convex set.

For the case when  $(\eta, \xi)$  has a multivariate normal distribution, we have derived an equivalent formulation for  $G(x) \leq \beta$  (first line in (2.59)). Slightly reformulated, this constraint is

$$-\Phi^{-1}(\beta) \|D^T x - d\| + \mu^T x \leq \mu_{n+1} \tag{2.79}$$

which is quite similar to (2.78). The sole difference is the different multiplier for the term  $\|D^T x - d\|$ . Taking  $\beta = 0.01$ , for example, we have  $\beta^{-\frac{1}{2}} = 10$  and  $-\Phi^{-1}(\beta) \approx 2.32$ . Thus, in the normally distributed case, requiring (2.78) instead of (2.79), a much stronger inequality results. Consequently, the feasible domain becomes much smaller in general. A prototype substitute problem takes the form

$$\left. \begin{aligned}
&\min c^T x \\
&\text{s.t. } \beta^{-\frac{1}{2}} \|D^T x - d\| + \mu^T x \leq \mu_{n+1} \\
&\quad x \in \mathcal{B}.
\end{aligned} \right\} \tag{2.80}$$

If for a given distribution, like the multivariate normal or the Cauchy distribution, an algebraic equivalent formulation exists, it makes no sense to use the stronger inequality (2.78). If, however, the distribution belongs to a class of distributions for which no equivalent algebraic formulation is known, or we have incomplete information regarding the distribution but have good estimates for the expected value and the covariance matrix, the substitute constraint (2.78) may provide a valuable modeling alternative. Notice that for *any distribution* with existing second moments, employing (2.78) in the model ensures that for the solution  $x^*$  the true inequality  $G(x^*) \leq \beta$  holds also. In other words, employing (2.80) is a conservative approach, which might be quite acceptable if, for instance,  $\beta$  represents the ruin probability of a company. Nevertheless, it may happen that the optimal objective value in (2.80)

becomes too high (too high costs, for instance), due to the narrower feasible domain in comparison with the feasible domain according to the true constraint  $G(x) \leq \beta$ .

Analogously, if  $G(x)$  is to be minimized in an SLP model, one might consider a substitute model with the upper bound from (2.76) in the objective. Thus, instead of

$$\left. \begin{array}{ll} \min & G(x) \\ \text{s.t.} & x \in \mathcal{B} \end{array} \right\}$$

we may consider the substitute problem

$$\left. \begin{array}{ll} \min & \frac{\|D^T x - d\|^2}{(-\mu^T x + \mu_{n+1})^2} \\ \text{s.t.} & x \in \mathcal{B}. \end{array} \right\}$$

Under the assumption that  $-\mu^T x + \mu_{n+1} > 0$  holds for all  $x \in \mathcal{B}$ , we get the equivalent formulation

$$\left. \begin{array}{ll} \min & \frac{\mu^T x - \mu_{n+1}}{\|D^T x - d\|} \\ \text{s.t.} & x \in \mathcal{B}, \end{array} \right\} \quad (2.81)$$

where equivalence means that the set of optimal solution of the two problems coincide. According to Proposition 2.32. on page 83, the objective function in (2.81) is pseudo-convex over  $\mathcal{B}$ , thus (2.81) is a convex programming problem. A comparison with (2.64) shows that the substitute problem and the original problem are equivalent in the case of the non-degenerate multivariate normal distribution (notice that (2.64) corresponds to maximizing  $G$ ). In the general case, the optimal objective value of the substitute problem (2.81) provides an upper bound on the optimal objective value of the original problem. Taking again the interpretation of  $G(x)$  as ruin probability, for any optimal solution  $x^*$  of (2.81), the ruin probability  $G(x^*)$  will not exceed the optimal objective value of (2.81). Concerning applicability of this approach, similar comments apply as for (2.80).

We would like to emphasize that, in general, both (2.80) and (2.81) are substitutes for the corresponding original problems, in general they are not equivalent to the true problems. Finally let us point out that this approach has first been suggested by Roy [292] and is utilized in the safety-first approaches to portfolio optimization, see Elton et al. [85].

### 2.2.4 The independent case

In this section we consider the joint probability function

$$G(x) = \mathbb{P}_\xi(T(\xi)x \geq h(\xi)) = \mathbb{P}_\xi(t_i^T(\xi)x \geq h_i(\xi), i = 1, \dots, s),$$

where the components of the  $n$ -dimensional random vector  $t_i(\xi)$  are the elements of the  $i^{\text{th}}$  row of the  $(s \times n)$  random matrix  $T(\xi)$ . We will assume in this section throughout that  $s > 1$  holds.

Our basic assumption is that the random vectors

$$(t_i^T(\xi), h_i(\xi)), \quad i = 1, \dots, s$$

are stochastically independent. Models of this type have first been formulated and studied by Miller and Wagner [234].

The stochastic independence implies that the random vector  $\zeta(x, \xi)$ , with  $\zeta_i(x, \xi) = t_i^T(\xi)x - h_i(\xi)$ ,  $i = 1, \dots, s$ , has stochastically independent components. Consequently, the probability function can be written in the independent case as follows:

$$\begin{aligned} G(x) &= \mathbb{P}(\zeta(x, \xi) \geq 0) = \prod_{i=1}^s \mathbb{P}(\zeta_i(x, \xi) \geq 0) \\ &= \prod_{i=1}^s \mathbb{P}(t_i^T(\xi)x \geq h_i(\xi)). \end{aligned} \quad (2.82)$$

We observe, that the probability function  $G(x)$  is the product of probability functions of the type, which have been studied in Section 2.2.3 on separate constraints; each term in the product involves a single random inequality.

Let us discuss the case first, when  $t_i(\xi) \equiv t_i \forall i$  holds, that means, we assume that only the right-hand-side is stochastic. Setting  $h(\xi) := \xi$ , we have

$$\begin{aligned} G(x) &= \mathbb{P}(t_i^T x \geq \xi_i, \quad i = 1, \dots, s) \\ &= F_{\xi_1, \dots, \xi_s}(t_1^T x, \dots, t_s^T x) \\ &= \prod_{i=1}^s F_{\xi_i}(t_i^T x). \end{aligned} \quad (2.83)$$

Distribution functions being monotonously increasing, the terms of the product are quasi-concave functions. This does not imply, however, the quasi-concavity of the product. Assuming positivity of the distribution functions, a natural idea is to transform the product into a sum, by a logarithmic transformation. The logarithm-function being strictly monotonically increasing, this would be suitable also from the optimization point of view. This way we get:

$$\log G(x) = \sum_{i=1}^s \log F_{\xi_i}(t_i^T x).$$

$\log G(x)$  will be concave, if the univariate distribution functions  $F_{\xi_i}$  are logconcave. As already noted by Miller and Wagner [234], log-concavity of univariate distribution functions is a thoroughly studied subject in statistics, more closely in reliability theory. It has been found that many important distributions, including the normal distribution, have logconcave distribution functions. For a recent summary see, for instance, Sengupta and Nanda [303] and the references therein.

Let us assume that the distribution functions  $F_{\xi_i}$  are logconcave  $\forall i$ , in the sense of the general Definition 2.13. on page 84.  $G(x)$ , being the product of logconcave functions, is logconcave (see page 85). Consequently, the probability constraint

$$G(x) \geq \alpha$$

defines a convex set,  $\forall \alpha \in [0, 1]$ . If the distribution functions are positive, the constraint can also be written as

$$\sum_{i=1}^s \log F_{\xi_i}(t_i^T x) \geq \log \alpha \quad (2.84)$$

for all  $\alpha \in (0, 1]$ .

If we drop the assumption of stochastic independence, but keep the supposition that only the right-hand-side is stochastic, then from (2.83) we see, that for the logconcavity of  $G$  it is sufficient, that the joint distribution function  $F_{\xi_1, \dots, \xi_s}$  is logconcave. This is true for several important distributions, and will be the subject of the subsequent Section 2.2.5.

Finally we discuss the situation under the stochastic independence assumption and random coefficients in the inequalities, see (2.82). We assume that the joint distributions of the rows are non-degenerate multivariate normal. For the separate terms of the product we can use the explicit form (2.57), derived in the section on separate probability constraints, thus resulting in:

$$G(x) = \prod_{i=1}^s \Phi \left( \frac{\mu^{(i)T} x - \mu_{n+1}^{(i)}}{\|D^{(i)} x - d^{(i)}\|} \right), \quad (2.85)$$

where  $\mu^{(i)}$ ,  $D^{(i)}$ , and  $d^{(i)}$  are the parameters of the normal distribution corresponding to the  $i^{\text{th}}$  row,  $\forall i$ . According to Proposition 2.36. on page 104, the terms of the product in (2.85) are pseudo-concave functions, at least on appropriate half-spaces. Unfortunately, this does not even imply that  $G(x)$  is quasi-concave. To ensure the convexity of  $\{x \mid G(x) \geq \alpha\}$  quite strong additional assumptions are needed. This topic will be further pursued in Section 2.2.6.

### 2.2.5 Joint constraints: random right-hand-side

In this section we consider a single probability constraint under the assumption that  $T(\xi) \equiv T$  holds, that means, we assume that the technology matrix is deterministic. We also simplify the notation by setting  $h(\xi) := \xi$ . Consequently, the probability constraint has the following form:

$$G(x) := \mathbb{P}_{\xi}(Tx \geq \xi) \geq \alpha \quad (2.86)$$

where  $T$  is an  $(s \times n)$  matrix and  $\xi$  is an  $s$ -dimensional random vector. Employing the probability distribution function  $F_\xi$ ,  $G(x)$  can be formulated as  $G(x) = F_\xi(Tx)$ . An alternative formulation for the probability constraint above is the following:

$$\begin{aligned} F_\xi(y) &\geq \alpha \\ y - Tx &= 0. \end{aligned} \tag{2.87}$$

From these representations it is clear, that for the convexity of the feasible domain

$$\mathcal{B}(\alpha) = \{x \mid G(x) \geq \alpha\}$$

it is sufficient, that the probability distribution function  $F_\xi$  is quasi-concave.

In the next subsection we will introduce the notion of generalized-concave probability measures. Via generalized-concavity properties of density functions this will lead to identifying several important classes of probability distributions for which  $F_\xi$  is quasi-concave. Subsequently we consider transformations which lead to generalized-concave probability functions. In the final subsection we consider SLP problems with joint probability functions in the objective.

### Generalized-concave probability measures

We will assume in this section that the probability distribution  $P_\xi$  is absolutely continuous (w.r. to the Lebesgue-measure), that means, we assume that the probability measure is generated by a probability density function. We will discuss various conditions concerning the probability measure  $P_\xi$  induced by  $\xi$ , under which the probability distribution function  $F_\xi$  is quasi-concave.

We begin by discussing generalized means, see Hardy, Littlewood, and Pólya [132].

Let  $a \geq 0$ ,  $b \geq 0$ , and  $\lambda \in [0, 1]$ . The generalized means  $\mathcal{M}_\gamma^\lambda(a, b)$  are defined as follows: for  $ab = 0$  let  $\mathcal{M}_\gamma^\lambda(a, b) = 0$ , for all  $\gamma \in \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$ . Otherwise, that is, if  $ab > 0$  holds, we define

$$\mathcal{M}_\gamma^\lambda(a, b) = \begin{cases} [\lambda a^\gamma + (1 - \lambda)b^\gamma]^{\frac{1}{\gamma}}, & \text{if } -\infty < \gamma < \infty \\ & \text{and } \gamma \neq 0 \\ a^\lambda b^{1-\lambda}, & \text{if } \gamma = 0 \\ \min\{a, b\}, & \text{if } \gamma = -\infty \\ \max\{a, b\}, & \text{if } \gamma = \infty. \end{cases} \tag{2.88}$$

The following monotonicity property of these generalized means will be used, see [132]:

$$\gamma_1 < \gamma_2 \implies \mathcal{M}_{\gamma_1}^\lambda(a, b) \leq \mathcal{M}_{\gamma_2}^\lambda(a, b), \quad \forall a, b \geq 0$$

with the inequality being strict, unless  $a = b$  or  $ab = 0$ . Based on these generalized means we define:

**Definition 2.16.** A nonnegative function  $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$  will be called  $\gamma$ -concave, if for any  $x, y \in \mathbb{R}^n$  and  $\lambda \in [0, 1]$  the following inequality holds:

$$f(\lambda x + (1 - \lambda)y) \geq \mathcal{M}_\gamma^\lambda(f(x), f(y)).$$

Let us note that in the literature this kind of generalized concave functions, as well as the generalized concave measures introduced later in this section, are usually called  $\alpha$ -concave, see for instance, Dancs and Uhrin [48] and Norkin and Roenko [245]. Because  $\alpha$  is used for probability levels in this chapter, we use the term  $\gamma$ -concave, instead.

Let  $f$  be a  $\gamma$ -concave function and  $\mathcal{C}^+ := \{x \mid f(x) > 0\}$ . The  $\gamma$ -concavity immediately implies that  $\mathcal{C}^+$  is a convex set. As already discussed for the logconcave case (c.f. Proposition 2.34. on page 85), this observation leads to the following alternative characterization: the nonnegative function  $f$  is  $\gamma$ -concave, if and only if  $\mathcal{C}^+$  is a convex set and the inequality in Definition 2.16. holds for all  $x, y \in \mathcal{C}^+$ .

For various  $\gamma$  values,  $\gamma$ -concavity can be interpreted over  $\mathcal{C}^+$  as follows (see the definition of the generalized means):

- $\gamma = +\infty$ :  $f$  is constant;
- $0 < \gamma < +\infty$ :  $f^\gamma$  is a concave function, note that  $\gamma = 1$  corresponds to ordinary concavity;
- $\gamma = 0$ :  $f$  is logconcave, that means,  $\log f$  is concave;
- $-\infty < \gamma < 0$ :  $f^\gamma$  is a convex function;
- $\gamma = -\infty$ :  $f$  is quasi-concave.

Notice that we have stated the properties only over  $\mathcal{C}^+$ . To see the reason, let us discuss the case  $\gamma = 1$ . A nonnegative function  $f$  is 1-concave, if it is concave over the convex set  $\mathcal{C}^+$ , where it is positive. If  $f$  is defined over  $\mathbb{R}^n$ , this does not mean that  $f$  is a concave function there. The following nonnegative function  $g : \mathbb{R} \rightarrow \mathbb{R}_+$

$$g(x) = \begin{cases} 1 - x^2 & \text{if } x \in [-1, 1] \\ 0 & \text{if } x \in (-\infty, -1) \text{ or } x \in (1, \infty) \end{cases}$$

is obviously 1-concave but it is not concave. Considering the well-known properties of concave functions, some caution is needed when 1-concave functions are dealt with. For instance, let  $g : \mathbb{R} \rightarrow \mathbb{R}_+$  and  $h : \mathbb{R} \rightarrow \mathbb{R}_+$  be both 1-concave functions, with  $\mathcal{C}_g^+ := \{x \mid g(x) > 0\}$  and  $\mathcal{C}_h^+ := \{x \mid h(x) > 0\}$ . Then for  $g + h$  we have  $\mathcal{C}_{g+h}^+ := \{x \mid g(x) + h(x) > 0\} = \mathcal{C}_g^+ \cup \mathcal{C}_h^+$ , which is a non-convex set in general. Thus, the sum of 1-concave functions is not necessarily 1-concave.

The monotonicity property of the generalized means implies: if  $f$  is  $\gamma_2$ -concave, then it is  $\gamma_1$ -concave, for all  $\gamma_1 < \gamma_2$ . In particular, if  $f$  is  $\gamma$ -concave for any  $\gamma \in [-\infty, \infty]$  then  $f$  is quasi-concave. For the implications concerning the various types of generalized concavity see Figure 2.8 on page 121.

Although pseudo-concavity does not fit into the class of  $\gamma$ -concave functions, logconcave functions, which are continuously differentiable over their domain of positivity, are also pseudo-concave there, see Proposition 2.35. on page 86. Consequently, for  $\gamma \geq 0$  the  $\gamma$ -concave functions, having the above smoothness property, are also pseudo-concave over their positivity domain.

We wish to extend the notion of  $\gamma$ -concavity to probability measures. For this we have to specify first, how a linear combination of sets should be defined. Let  $A$  and  $B$  two subsets of  $\mathbb{R}^r$  and let  $\lambda \in \mathbb{R}$ . We employ the following definitions:

$$\begin{aligned} A + B &= \{x \mid \exists y \in A \text{ and } \exists z \in B, \text{ such that } x = y + z\}, \\ \lambda A &= \{x \mid \exists y \in A \text{ such that } x = \lambda y\}. \end{aligned} \quad (2.89)$$

Figure 2.7 shows the convex combination of two sets. For the properties of these operations on sets see, for instance, Rockafellar [281].

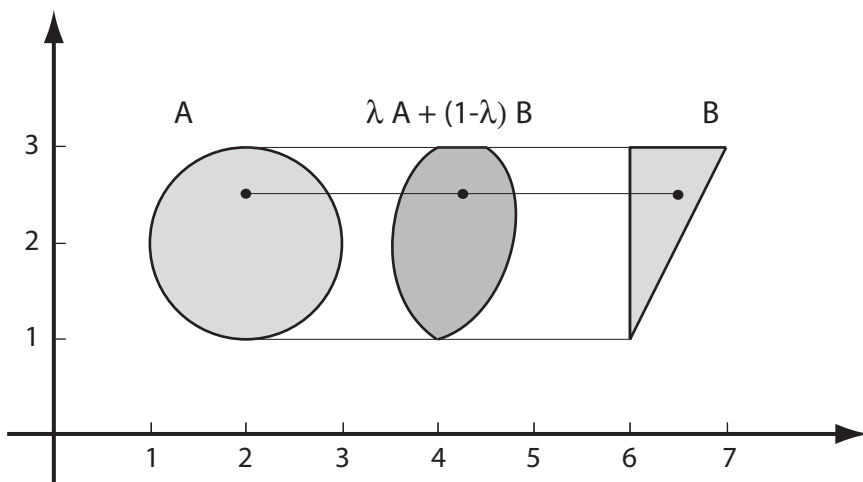


Fig. 2.7 Convex combination of two sets with  $\lambda = \frac{1}{2}$ .

We will confine ourselves to the case, when both sets are convex. Let  $A$  and  $B$  be convex sets;  $\lambda, \mu \in \mathbb{R}$ . The following properties are important for the future discussion:

- $A + B$  and  $\lambda A$  are convex sets, see [281].
- Let  $\lambda \geq 0$  and  $\mu \geq 0$ . Then  $(\lambda + \mu)A = \lambda A + \mu A$  (without the convexity of  $A$  only  $(\lambda + \mu)A \subset \lambda A + \mu A$  holds). See [281].
- If either  $A$  or  $B$  is open, then  $A + B$  is open.
- If both  $A$  and  $B$  are closed, and at least one of them is bounded, then  $A + B$  is closed. The sum of two unbounded closed convex sets need not to be closed, see [281]. If both  $A$  and  $B$  are closed then  $A + B$  is Borel-measurable, see, for instance, [88].



- If  $A$  is convex, then it is obviously Lebesgue-measurable, because the boundary has Lebesgue-measure 0.
- If  $A$  is convex, then it is not necessarily Borel-measurable. To see this, let us construct a convex set in  $\mathbb{R}^2$  as follows: Let us take a non-Borel-measurable set  $\mathcal{K}$  on the interval  $[0, 2\pi)$  (for the existence of such a set see, for instance, Billingsley [21]) and let us map this set onto the boundary of the open unit disc in  $\mathbb{R}^2$  by the mapping  $\Psi : \mathcal{K} \rightarrow \mathbb{R}^2, x \rightarrow (\cos x, \sin x)$ . The union of the open unit disc and the image of  $\mathcal{K}$  under  $\Psi$  is obviously convex, and, as a union of a Borel-measurable set (the open unit disc), and a non-Borel-measurable set, it cannot be Borel-measurable.
- The sum of two Borel-measurable sets is not necessarily Borel-measurable, see Erdős and Stone [88].

As a next step, we will define generalized concavity properties of probability measures, in analogy with Definition 2.16. Considering the list of properties above, one must be careful in working with convex combinations of Borel-sets. Therefore we formulate the definition as follows:

**Definition 2.17.** *The probability measure  $\mathbb{P}$  on the Borel-sets  $\mathbb{B}^r$  is called  $\gamma$ -concave, if for any convex, measurable sets  $A$  and  $B$  and any  $\lambda \in [0, 1]$ , for which  $\lambda A + (1 - \lambda)B$  is Borel-measurable, the following inequality holds:*

$$\mathbb{P}(\lambda A + (1 - \lambda)B) \geq \mathcal{M}_\gamma^\lambda(\mathbb{P}(A), \mathbb{P}(B)).$$

A  $\gamma$ -concave probability measure with  $\gamma = -\infty$  will be called a *quasi-concave*. In this case the defining inequality takes the form

$$\mathbb{P}(\lambda A + (1 - \lambda)B) \geq \min\{\mathbb{P}(A), \mathbb{P}(B)\}.$$

For  $\gamma = 0$  we have a *logconcave* probability measure, with the defining inequality

$$\mathbb{P}(\lambda A + (1 - \lambda)B) \geq \mathbb{P}(A)^\lambda \mathbb{P}(B)^{1-\lambda}.$$

Let  $\xi$  be a random variable and  $\mathbb{P}_\xi$  the induced measure on  $\mathbb{B}^r$ . We denote by  $F_\xi$  the probability distribution function of  $\xi$ . For any convex, closed set  $A$  in  $\mathbb{R}^r$  let us introduce the function  $\Gamma_A(y) = \mathbb{P}_\xi(A + \{y\})$ . Then the following proposition holds.

**Proposition 2.37.** *If  $\mathbb{P}_\xi$  is a  $\gamma$ -concave measure, then  $\Gamma_A$  is a  $\gamma$ -concave function.*

*Proof:* Let  $x, y \in \mathbb{R}^r, \lambda \in [0, 1]$ . Then we have

$$\begin{aligned} \Gamma_A(\lambda x + (1 - \lambda)y) &= \mathbb{P}_\xi(A + \{\lambda x + (1 - \lambda)y\}) \\ &= \mathbb{P}_\xi([\lambda + (1 - \lambda)]A + \{\lambda x + (1 - \lambda)y\}) \\ &= \mathbb{P}_\xi(\lambda[A + \{x\}] + (1 - \lambda)[A + \{y\}]) \\ &\geq \mathcal{M}_\gamma^\lambda(\mathbb{P}(A + \{x\}), \mathbb{P}(A + \{y\})) \\ &= \mathcal{M}_\gamma^\lambda(\Gamma_A(x), \Gamma_A(y)). \end{aligned}$$

□

Let us assume that  $\mathbb{P}_\xi$  is  $\gamma$ -concave. Taking  $A = \mathbb{R}_-^r$  we get from Proposition 2.37., that  $F_\xi$  is  $\gamma$ -concave. Consequently,  $\{x \mid \mathbb{P}_\xi(Tx \geq \xi) \geq \alpha\}$  is a convex set,  $\forall \alpha \in [0, 1]$  (see (2.86)).

Let us consider

$$H(y) := \mathbb{P}_\xi\{y \mid \xi \geq y\}.$$

Choosing now  $A = \mathbb{R}_+^r$ , Proposition 2.37. implies, that  $H$  is also  $\gamma$ -concave. Consequently,  $\{x \mid \mathbb{P}_\xi(Tx \leq \xi) \geq \alpha\}$  is also a convex set,  $\forall \alpha \in [0, 1]$ .

The above considerations imply, that for showing that the distribution function  $F$  is  $\gamma$ -concave, it is sufficient to prove the  $\gamma$ -concavity of the probability distribution  $\mathbb{P}_\xi$ .

The following fundamental theorem links, for continuous distributions, the  $\gamma$ -concavity of the probability density function with the  $\gamma$ -concavity of  $\mathbb{P}_\xi$ .

**Theorem 2.8.** *Let  $f$  be a  $\gamma$ -concave probability density function for the probability distribution of the  $r$ -dimensional random variable  $\xi$ . Let  $-\frac{1}{r} \leq \gamma \leq \infty$ . Then  $\mathbb{P}_\xi$  is an  $\frac{\gamma}{1+\gamma}$ -concave probability measure.*

*Proof:* Let  $\lambda \in [0, 1]$  and assume that the convex sets  $A, B$ , and  $\lambda A + (1 - \lambda)B$  are Borel-measurable. The  $\gamma$ -concavity of  $f$  implies:

$$\begin{aligned} \mathbb{P}_\xi(\lambda A + (1 - \lambda)B) &= \int_{\lambda A + (1 - \lambda)B} f(z) dz \\ &\geq \int_{\lambda A + (1 - \lambda)B} \left( \sup_{\lambda x + (1 - \lambda)y = z} \mathcal{M}_\gamma^\lambda(f(x), f(y)) \right) dz. \end{aligned}$$

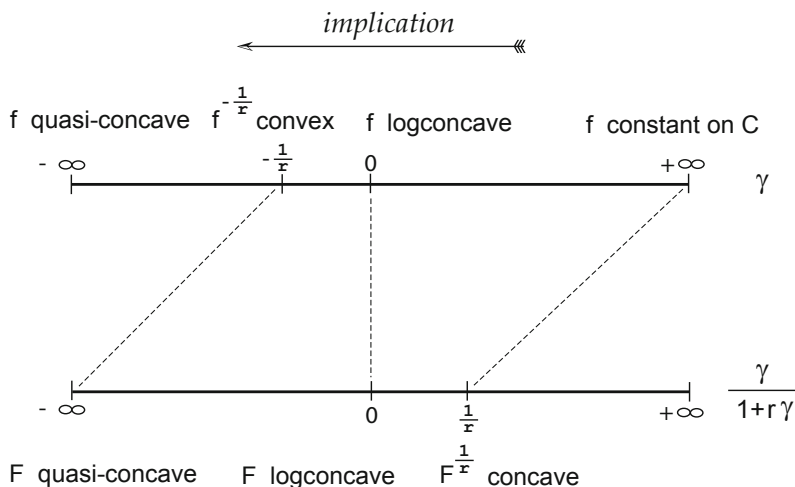
Now we apply an integral-inequality, see Prékopa [266] ( for  $\gamma = 0$  it is called Prékopa's inequality):

$$\begin{aligned} \mathbb{P}_\xi(\lambda A + (1 - \lambda)B) &\geq \int_{\lambda A + (1 - \lambda)B} \left( \sup_{\lambda x + (1 - \lambda)y = z} \mathcal{M}_\gamma^\lambda(f(x), f(y)) \right) dz \\ &\geq \mathcal{M}_{\frac{\gamma}{1+\gamma}}^\lambda \left( \int_A f(x) dx, \int_B f(y) dy \right) \\ &= \mathcal{M}_{\frac{\gamma}{1+\gamma}}^\lambda (\mathbb{P}_\xi(A), \mathbb{P}_\xi(B)) \end{aligned}$$

which completes the proof.  $\square$

For some ranges of  $\gamma$ -values we summarize the assertion of the theorem, together with the implications from Theorem 2.37., see also Figure 2.8. For this let  $C := \{x \mid f(x) > 0\}$  and let us assume that  $C$  is a convex set.

- $f$  is constant over  $C \implies F^{\frac{1}{r}}$  and  $H^{\frac{1}{r}}$  are concave, consequently both  $F$  and  $H$  are logconcave and therefore also quasi-concave.
- $f$  is logconcave  $\implies F$  and  $H$  are logconcave and therefore quasi-concave, too.
- $f^{-\frac{1}{r}}$  is convex  $\implies F$  and  $H$  are quasi-concave.



**Fig. 2.8**  $\gamma$ -concave density functions versus  $\frac{\gamma}{1+r\gamma}$ -concave distribution functions.

Logconcave functions have several nice properties. We will need the following fact:

**Theorem 2.9.** *Prékopa [259]. Let  $f : \mathbb{R}^{n+m} \rightarrow \mathbb{R}_+$  be a logconcave function. Then*

$$g(x) := \int_{\mathbb{R}^m} f(x, y) dy$$

*is a logconcave function on  $\mathbb{R}^n$ .*

*Proof:* See Prékopa [266]. □

If  $f$  is a logconcave density function then this theorem implies that all marginal density functions are logconcave, too.

If  $f$  and  $g$  are two logconcave density functions on  $\mathbb{R}^n$  then their convolution is also logconcave. In fact, the logconcavity of  $f$  implies that  $h(x, y) := f(x - y)$  is logconcave in  $\mathbb{R}^{2n}$  (see the remark on page 85). Thus  $f(x - y)g(y)$  is logconcave in  $\mathbb{R}^{2n}$ . Applying Theorem 2.9. yields the result.

For  $\gamma = 0$  Theorem 2.8. has first been established by Prékopa in 1971 [258], by Leindler 1972 [203], and in its final form by Prékopa 1973 [259]. Dinghas 1957 [72] proved the theorem for  $\gamma > 0$ . Borell proved the theorem in full generality in 1975 [31].

The breakthrough in the field of generalized concave measures and their application in stochastic programming has been achieved by Prékopa, who developed the theory of logarithmic concave probability measures. These fundamental results have inspired several authors: papers with alternative proofs have appeared, the theory has been extended to quasi-concave measures, and applications in stochastic programming, statistics, and economics have been studied. For a comprehensive discussion

of these results see Prékopa [266] and the references therein. Here we confine ourselves to refer to Brascamp and Lieb [33], Dancs and Uhrin [48], Das Gupta [54], Kallberg and Ziemba [173], Norkin and Roenko [245], and Rinott [277]. Converse results have been obtained, for instance, by Borell [31], Brascamp and Lieb [33], and Kall [154].

As applications of Theorem 2.8., below we give some examples for multivariate probability distributions, for which the probability distribution function is quasi-concave or even logconcave. The probability distribution- and density functions will be denoted by  $F$  and  $f$ , respectively. For a square matrix  $D$ , its determinant will be denoted by  $|D|$ . For multivariate distributions and their usage in statistics see, for instance, Johnson and Kotz [149] and Mardia, Kent, and Bibby [216].

- *Uniform distribution on a convex set.* The density function is

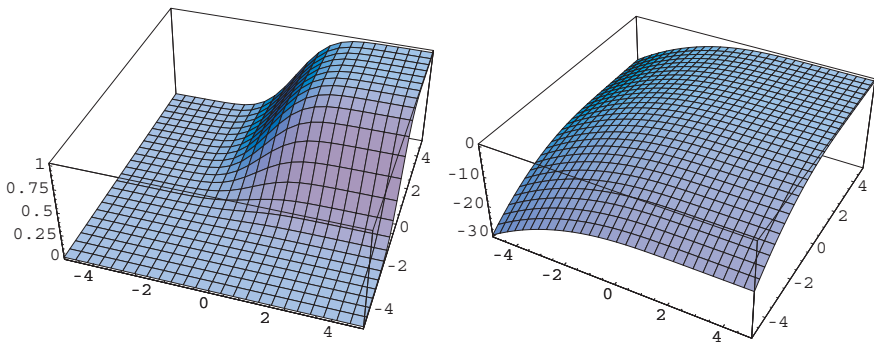
$$f(x) = \begin{cases} \frac{1}{\lambda(C)} & \text{if } x \in C \\ 0 & \text{otherwise,} \end{cases}$$

where  $C \subset \mathbb{R}^s$  is a bounded convex set with a positive Lebesgue-measure  $\lambda(C)$ .  $f$  is obviously logconcave thus  $F$  is logconcave, too.

- *Non-degenerate normal distribution.* The density function of this distribution is positive on  $\mathbb{R}^r$  and is given in (2.51) on page 99. Taking logarithm and neglecting the additive constant results in

$$-\frac{1}{2}(y - \mu)^T \Sigma^{-1}(y - \mu).$$

This is a concave function, because with  $\Sigma$ ,  $\Sigma^{-1}$  is also positive definite, see for instance, Horn and Johnson [141]. Thus  $f$  is logconcave implying the logconcavity of  $F$ . Figure 2.9 shows the standard bivariate normal distribution function and its logarithm.



**Fig. 2.9** The bivariate standard normal distribution function and its logarithm.

- *Dirichlet distribution.* This is the joint distribution of the random variables  $\eta_j = \xi_j \left[ \sum_{i=0}^r \xi_i \right]^{-1}$ ,  $j = 1, \dots, r$ , where  $\xi_j$ ,  $j = 0, \dots, r$  are independent random variables,  $\xi_j$  having  $\chi^2$ -distribution with  $v_j > 0$  degrees of freedom. The density function of this distribution is

$$f(x) = \begin{cases} \frac{\Gamma(\sum_{j=0}^s \theta_j)}{\prod_{j=0}^s \Gamma(\theta_j)} (1 - \sum_{j=1}^s x_j)^{\theta_0-1} \prod_{j=1}^s x_j^{\theta_j-1}, & \text{if } x > 0, \sum_{j=1}^s x_j < 1, \\ 0, & \text{otherwise,} \end{cases}$$

where  $\theta_j$ ,  $j = 0, \dots, r$  are the parameters of the distribution;  $\theta_j = \frac{1}{2}v_j \forall j$ . On the convex set  $\mathcal{C}^+ := \{x \mid x > 0, \sum_{j=1}^s x_j < 1\}$  the density function is positive and it is zero if  $x \notin \mathcal{C}^+$ . Therefore, see Proposition 2.34. on page 85, for checking logconcavity, it is sufficient to consider  $\log f(x)$  over  $\mathcal{C}^+$ . Apart of an additive constant, we have for  $x \in \mathcal{C}^+$ :

$$\log f(x) = (\theta_0 - 1) \log(1 - \sum_{j=1}^s x_j) + \sum_{j=1}^s (\theta_j - 1) \log x_j.$$

If  $\theta_j \geq 1 \forall j$  then this is a linear combination, with nonnegative coefficients, of concave functions, therefore  $\log f(x)$  is concave on  $\mathcal{C}^+$ . Let us remark that the concavity of the first term in the right-hand-side follows from the fact, that substitution of an affine-linear function into a concave function preserves concavity. We have got: provided that  $\theta_j \geq 1$  for  $j = 0, \dots, s$  holds,  $f$  is logarithmic concave implying the logconcavity of  $F$ .

- *Wishart distribution.* This is the joint distribution of the elements of the sample covariance matrix for a multivariate normal population. Let us consider a sample with sample-size  $N > s$  from a population consisting of  $s$ -dimensional random vectors having a multivariate normal distribution with covariance matrix  $C$ . The density function for this distribution is the following:

$$f(X) = \begin{cases} \gamma |X|^{\frac{1}{2}(N-s-2)} e^{-\frac{1}{2}\text{Tr}(C^{-1}X)}, & \text{if } X \text{ is positive definite} \\ 0, & \text{otherwise,} \end{cases}$$

where  $X$  is an  $(s \times s)$  symmetric matrix and

$$\gamma = |C|^{-\frac{N-1}{2}} \left[ 2^{\frac{(N-1)s}{2}} \pi^{\frac{s(s-1)}{4}} \prod_{j=1}^s \Gamma\left(\frac{N-j}{2}\right) \right]^{-1}$$

holds. For an  $(s \times s)$  matrix  $D$ ,  $\text{Tr} D := \sum_{j=1}^s D_{jj}$  denotes the *trace* of  $D$ . We wish to check whether  $f$  is logconcave. For this we first observe that the set

$\mathcal{C}^+ := \{X \mid X \text{ is symmetric positive definite}\}$  is obviously a convex subset of the linear space of symmetric  $(s \times s)$  matrices. Therefore it is sufficient to consider  $\log f$  on  $C^+$ :

$$\log f(X) = \log \gamma + \frac{1}{2}(N - s - 2) \log |X| - \frac{1}{2} \text{Tr} C^{-1} X.$$

The third term is obviously linear in  $X$ . According to an inequality of Fan (see, for instance, Beckenbach and Bellman [13]), the function  $|X|$  is a logconcave function of  $X$ . Therefore, if  $N \geq s - 2$  then  $f$  is logconcave and so  $F$  is logconcave, too.

- *t-distribution (Student-distribution).* We consider the joint distribution of  $\eta_j = \xi_j \left( \frac{\zeta}{\sqrt{\nu}} \right)^{-1}$ ,  $j = 1, \dots, r$ , where  $(\xi_1, \dots, \xi_r)$  has a joint standardized non-degenerate multivariate normal distribution with correlation matrix  $R$ .  $\zeta$  has a  $\chi$ -distribution with  $\nu$  degrees of freedom. The density function for this distribution is positive on  $\mathbb{R}^s$  and has the analytical form

$$f(x) = \frac{\Gamma(\frac{1}{2}(\nu + s))}{(\pi\nu)^{\frac{s}{2}} \Gamma(\frac{\nu}{2}) |R|^{\frac{1}{2}}} \left(1 + \frac{1}{\nu} x^T R^{-1} x\right)^{-\frac{1}{2}(\nu + s)},$$

where the parameters are  $R$ , a symmetric positive definite matrix, and the positive integer  $\nu$ , interpreted as degrees of freedom.  $f^{-\frac{1}{s}}$  is, apart of a positive multiplicative constant, as follows:

$$g(x) := \left(1 + \frac{1}{\nu} x^T R^{-1} x\right)^{\frac{1}{2}(1 + \frac{\nu}{s})},$$

which is a convex function on  $\mathbb{R}^s$ . To see this, let us remark first that

$$h(x) := \left(1 + \frac{1}{\nu} x^T R^{-1} x\right)^{\frac{1}{2}} = \left( (x^T, 1) \begin{pmatrix} R^{-1} & 0 \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} \right)^{\frac{1}{2}}$$

is convex because the positive definite matrix above induces a norm in  $\mathbb{R}^{s+1}$ . We have  $g = h^{1 + \frac{\nu}{s}}$ , therefore the convexity of  $g$  follows from the fact, that substituting a convex function into a monotonically increasing convex function results in a convex composite function. Thus  $f^{-\frac{1}{s}}$  is convex, implying that  $F$  is quasi-concave.

- *Univariate gamma distribution.* The density function of this distribution is

$$f(x) = \begin{cases} \frac{\lambda^\vartheta}{\Gamma(\vartheta)} x^{\vartheta-1} e^{-\lambda x}, & \text{if } x > 0 \\ 0, & \text{otherwise,} \end{cases}$$

where  $\lambda > 0$  and  $\vartheta > 0$  are parameters. This distribution will be denoted by  $\mathcal{G}(\lambda, \vartheta)$ . If  $\lambda = 1$ , then the distribution is called a standard gamma distribution. Assuming  $x > 0$  and taking logarithm we observe that  $f$  is logconcave, provided

that  $\vartheta \geq 1$  holds. If an  $s$ -dimensional random vector  $\eta$  has stochastically independent components  $\eta_i \sim \mathcal{G}(\lambda_i, \vartheta_i)$  and  $\vartheta_i \geq 1 \forall i$  holds, then  $\eta$  has a logconcave distribution. This follows by considering the density function of  $\eta$ , which is the product of the one-dimensional density functions of the components. The univariate densities being logconcave, their product is logconcave, too.

### Generalized-concave distribution functions

So far we have discussed one way for ensuring generalized concavity of the distribution function  $F_\xi$ . The method, applicable for continuous distributions, has been the following: the generalized concavity of the probability density function has been studied, which implied via Theorem 2.8. the generalized concavity of  $F_\xi$ . For several important multivariate distributions it turned out that  $F_\xi$  is pseudo-concave, or that they even have the more important logconcavity property.

Another possibility has been discussed in Section 2.2.4. Under the assumption that the components of  $\xi$  are stochastically independent, the joint distribution function is the product of the one-dimensional marginal distribution functions, that means,

$$F_\xi(y) = \prod_{i=1}^s F_{\xi_i}(y_i).$$

If the marginal distribution functions  $F_{\xi_i}$  are logconcave then  $F_\xi$  will be logconcave, too.

In the sequel we explore further ways for ensuring generalized concavity properties of the probability distribution function. The idea is to apply transformations to random vectors having generalized-concave distributions, in order to obtain distributions for which the probability distribution function again has some generalized concavity properties.

The subsequent theorems and the insight behind their application in stochastic programming have been first found by Prékopa for the logconcave case. Their extension to the  $\gamma$ -concave case is straightforward.

The following general theorem gives a sufficient condition for generalized concavity of composite functions. See, for instance, Prékopa [266] and for an extension Tamm [324].

We consider the following probability function:

$$M(x) = \mathbb{P}_\xi \{g_i(x, \xi) \geq 0, i = 1, \dots, s\} = \mathbb{P}_\xi \{g(x, \xi) \geq 0\}, \quad (2.90)$$

where  $g_i: \mathbb{R}^n \times \mathbb{R}^r \rightarrow \mathbb{R}$ ,  $i = 1, \dots, s$ ;  $g^T(x, \xi) = (g_1^T(x, \xi), \dots, g_s^T(x, \xi))$ .

**Theorem 2.10.** *Let  $g_i$ ,  $i = 1, \dots, m$  be quasi-concave functions, that means, let  $g_i(\cdot, \cdot)$  be jointly quasi-concave in both arguments. For the sake of simplicity we also assume that  $g$  is continuous. Assume further, that  $\xi$  has a  $\gamma$ -concave probability distribution. Then  $M(x)$  is a  $\gamma$ -concave function.*

*Proof:* Let  $\mathcal{H}(x) := \{z \mid g(x, z) \geq 0\} \subset \mathbb{R}^s$ . Due to our assumptions these sets are convex and closed  $\forall x$  and we have  $M(x) = \mathbb{P}_\xi(\mathcal{H}(x))$ . Let  $\lambda \in (0, 1)$ ,  $x, y \in \mathbb{R}^n$ . The basic ingredient of the proof is the following inclusion, which can be proved in a straightforward way:

$$\mathcal{H}(\lambda x + (1 - \lambda)y) \supset \lambda \mathcal{H}(x) + (1 - \lambda)\mathcal{H}(y).$$

Using this and the  $\gamma$ -concavity of the probability measure, we immediately get:

$$\begin{aligned} M(\lambda x + (1 - \lambda)y) &= \mathbb{P}_\xi(\mathcal{H}(\lambda x + (1 - \lambda)y)) \\ &\geq \mathbb{P}_\xi(\lambda \mathcal{H}(x) + (1 - \lambda)\mathcal{H}(y)) \\ &\geq \mathcal{M}_\gamma^\lambda(\mathbb{P}_\xi(\mathcal{H}(x)), \mathbb{P}_\xi(\mathcal{H}(y))) \\ &= \mathcal{M}_\gamma^\lambda(M(x), M(y)). \end{aligned}$$

□

As an application of this theorem we will show, how it can be applied to prove logconcavity of the log-normal distribution function.

- *Log-normal distribution.* Let the random variables  $\xi_1, \dots, \xi_s$  have a joint non-degenerate multivariate normal distribution. The joint distribution of the random variables  $\eta_i = e^{\xi_i}$ ,  $i = 1, \dots, s$  is called a multivariate log-normal distribution. The density function of this distribution is not logconcave, see Prékopa [266]. For the joint distribution function  $F_\eta$  we have:

$$\begin{aligned} F_\eta(x_1, \dots, x_s) &= \mathbb{P}_\eta(\eta_1 \leq x_1, \dots, \eta_s \leq x_s) \\ &= \mathbb{P}_\xi(x_1 - e^{\xi_1} \geq 0, \dots, x_s - e^{\xi_s} \geq 0). \end{aligned}$$

In the preceding section we have seen that the probability measure of a non-degenerate multivariate normal distribution is logconcave. Theorem 2.10. can be applied with  $\gamma = 0$  thus showing that  $F_\eta$  is a logconcave function.

Let us consider next the effect of linear transformations of random variables having  $\gamma$ -concave distributions. The following theorem holds:

**Theorem 2.11.** *Let  $\xi$  be an  $s$ -dimensional random vector,  $D$  an  $(r \times s)$  matrix, and  $\zeta = D\xi + d$ . If  $\xi$  has a  $\gamma$ -concave distribution then the distribution of  $\zeta$  is also  $\gamma$ -concave.*

*Proof:* Let  $\lambda \in (0, 1)$  and let  $A, B, C_\lambda := \lambda A + (1 - \lambda)B$  be Borel-measurable convex sets in  $\mathbb{R}^r$ . Then their inverse images in  $\mathbb{R}^s$  under the affine linear transformation defined by  $D$  and  $d$ , that means,

$$\begin{aligned} \bar{A} &:= \{x \mid Dx + d \in A\}, \\ \bar{B} &:= \{x \mid Dx + d \in B\}, \text{ and} \\ \bar{C}_\lambda &:= \{x \mid Dx + d \in C_\lambda\} \end{aligned}$$



are Borel-measurable convex sets in  $\mathbb{R}^s$ . It is easy to see that

$$\bar{C}_\lambda \supset \lambda \bar{A} + (1 - \lambda) \bar{B}$$

holds. Using this we get

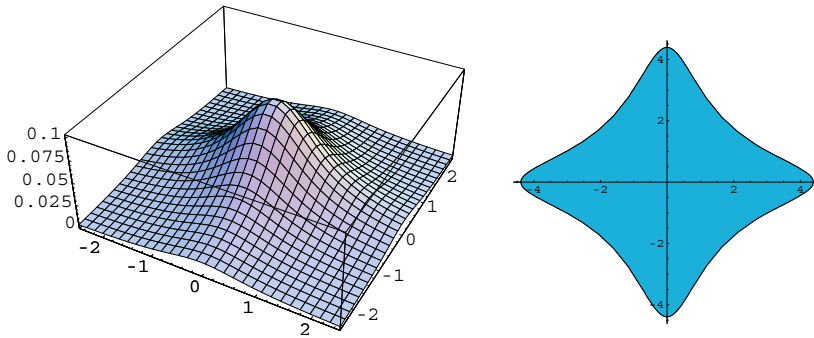
$$\begin{aligned} \mathbb{P}_\zeta(\lambda A + (1 - \lambda)B) &= \mathbb{P}_\xi(\bar{C}_\lambda) \\ &\geq \mathbb{P}_\xi(\lambda \bar{A} + (1 - \lambda) \bar{B}) \\ &\geq \mathcal{M}_\gamma^\lambda(\mathbb{P}_\xi(\bar{A}), \mathbb{P}_\xi(\bar{B})) \\ &= \mathcal{M}_\gamma^\lambda(\mathbb{P}_\zeta(A), \mathbb{P}_\zeta(B)). \end{aligned}$$

□

This theorem can be utilized to study generalized concavity properties of distributions, which are derived in a similar way, as the multivariate normal distribution. We take  $s$  stochastically independent continuous random variables, each of them having a  $\gamma$ -concave density function. The joint density function is then the product of the density functions of the components. If this joint density function is  $\gamma$ -concave, then via Theorem 2.11.,  $\zeta = D\xi + d$  will have a  $\gamma$ -concave distribution. Especially, if the components of  $\xi$  have logconcave densities ( $\gamma = 0$ ), then the joint density function of  $\xi$  will be logconcave (see page 85).

- *The multivariate normal distribution.* We consider the multivariate normal distribution, see Definition 2.14. on Page 98. In Section 2.2.5 we have proved, by applying Theorem 2.8., that the non-degenerate multivariate normal distribution is logconcave. Without the non-degeneracy assumption we can proceed as follows. Recall (Definition 2.14. on page 98) that the  $r$ -dimensional random vector  $\zeta$  has a multivariate normal distribution, if  $\zeta = B\xi + \mu$  holds, where  $B$  is an  $(r \times s)$  matrix,  $\mu \in \mathbb{R}^r$  holds, and the components of  $\xi$  are independent and have a standard normal distribution. The joint probability distribution of  $\xi$  is then obviously non-degenerate multivariate normal. Thus, Theorem 2.8. implies that  $\xi$  has a logconcave probability distribution. Consequently, the application of Theorem 2.11. yields the logconcavity of the probability distribution of  $\zeta$  and thus the logconcavity of the multivariate normal distribution in the general case.
- *A multivariate gamma distribution.* In the preceding section on Page 124 we have seen that the univariate gamma distribution has a logconcave density function, therefore our technique can be used in this case, too. Prékopa and Szántai [270] have defined a multivariate gamma distribution as follows. Let  $\xi$  be a  $s = 2^r - 1$  dimensional random vector with stochastically independent components. The components are assumed to have standard gamma distributions, see Page 124. Let  $D$  be an  $(r \times 2^r - 1)$  matrix with nonzero columns and components equal to 0 or 1. The distribution of  $\zeta := D\xi$  is called a multivariate gamma distribution. If for the parameter  $\vartheta \geq 1$  holds, then Theorem 2.11. implies that the distribution of  $\zeta$  is logconcave. If  $\vartheta < 1$  then the distribution of  $\zeta$  is not

necessarily logconcave, but the joint distribution function  $F_{\xi}$  is still a logconcave function, see [270].



**Fig. 2.10** The graph of the bivariate standard Cauchy-distribution and the upper level set corresponding to level 0.005.

In Section 2.2.3 we have considered a multivariate Cauchy distribution, which is derived on the basis of an affine linear transformation as discussed above, see Definition 2.15. on page 109. A natural idea is trying to apply Theorem 2.11. for deriving some  $\gamma$ -concavity property of the multivariate Cauchy distribution. Notice, however, that the density function of the univariate Cauchy distribution is not logconcave. Moreover, as it can easily be seen, the product of the density functions of standard univariate Cauchy distributions is not even quasi-concave, see Figure 2.10. Therefore, see Figure 2.8, the joint density function of  $\xi$  is not  $\gamma$ -concave, for any  $\gamma$ . Consequently, our technique does not go through in this case. Notice, however, that there are other generalizations of the Cauchy distribution to the multivariate case, where the distribution is quasi-concave, see Prékopa [268].

Finally let us comment on the case when  $\xi$  has a finite discrete distribution. Prékopa [266] gave a definition of logconcavity of such distributions and studied their properties. In Dentcheva et al. [65], the authors extend this notion to  $r$ -concave discrete distributions, where  $r$ -concavity corresponds to  $\gamma$ -concavity (see Section 2.2.5) in the continuous case, and is appropriately modified for the discretely distributed case. The authors also report on algorithmically relevant applications by providing bounds on the optimal objective value of SLP problems with probabilistic constraints.

## Maximizing joint probability functions

For the case when the probability function is in the objective, we formulate the prototype problem

$$\begin{aligned} \max \quad & G(x) \\ \text{s.t.} \quad & x \in \mathcal{B}, \end{aligned} \tag{2.91}$$

where  $G$  is the probability function  $G(x) = \mathbb{P}_\xi(Tx \geq \xi)$  and  $\mathcal{B}$  is a polyhedral set determined by linear inequalities and/or equalities, see (2.8) on page 74.

If  $G$  is logconcave and differentiable then it is also pseudo-concave, see Proposition 2.35. on page 86. This is the case, for instance, for the non-degenerate multivariate normal distribution or for the log-normal distribution (see pages 122 and 126). Consequently, for logconcave distributions, (2.91) is a linearly constrained convex optimization problem. Some other distributions only have the quasi-concavity property, like the multivariate  $t$ -distribution (see page 124). In such cases (2.91) has a quasi-concave objective function and the problem may have local maxima which are not global solutions; the problem becomes much more difficult to solve numerically.

Note that (2.91) has been formulated as a maximization problem. Assuming  $s > 1$ , that is, assuming that  $\xi$  is a random vector, this is the only way for arriving at convex programming problems. Reversing the random inequality does not help in this respect: with  $G$  the function  $\hat{G} : \hat{G}(x) = \mathbb{P}_\xi(Tx \leq \xi)$  is also logconcave, see the discussion on page 120. For reversing the random inequality in the multivariate case see also (2.31) on page 89. Thus, for  $\xi$  having a logconcave distribution and assuming  $s > 1$ , the counterpart of (2.91) involving minimization is a much more difficult problem numerically than (2.91).

### 2.2.6 Joint constraints: random technology matrix

In this section we consider the probability function in full generality

$$G(x) = \mathbb{P}_\xi(x \mid T(\xi)x \geq h(\xi)),$$

where the  $(s \times n)$  technology matrix  $T(\xi)$  is also allowed to be stochastic. In Section 2.2.3 on separate probability constraints we have assumed that  $s = 1$  holds. We have seen that the feasible domain is convex under various further assumptions concerning the probability distribution and the probability level  $\alpha$ . If  $s > 1$ , then the convexity of the feasible domain can only be ensured under quite strong assumptions. We will discuss the case, when the joint distribution of the random entries is multivariate normal.

The matrix of random entries  $(T(\xi), h(\xi))$  will be considered both column-wise and row-wise, therefore we introduce the notation:

$$(T(\xi), h(\xi)) = (\zeta^{(1)}, \dots, \zeta^{(n)}, \zeta^{(n+1)}) = \begin{pmatrix} \eta^{(1)T} \\ \vdots \\ \eta^{(s)T} \end{pmatrix}.$$

Here the  $s$ -dimensional random vector  $\zeta^{(j)}$  denotes the  $j^{\text{th}}$  column of  $T(\xi)$  for  $j \leq n$ , and the right-hand-side  $h(\xi)$  for  $j = n + 1$ .

Let  $\zeta^T = (\zeta^{(1)T}, \dots, \zeta^{(n+1)T})$  be the random vector consisting of all random entries in columns major order. The  $n+1$ -dimensional random vector  $\eta^{(i)}$  stands for the  $i^{\text{th}}$  row,  $1 \leq i \leq s$ , that is,  $\eta^{(i)T} = (T_{i1}(\xi), \dots, T_{in}(\xi), h_i(\xi))$ . Let  $\eta^T = (\eta^{(1)T}, \dots, \eta^{(s)T})$  be the random vector consisting of all random entries in rows major order.

For any vector  $x \in \mathbb{R}^n$  let  $\hat{x} \in \mathbb{R}^{n+1}$  be  $\hat{x}^T = (x_1, \dots, x_n, -1)$ . For simplicity of notation in this section we drop the explicit reference to  $\xi$  in  $\zeta(x, \xi)$ . We have the following alternative representations

$$\begin{aligned} \zeta(x) &:= \sum_{i=1}^n \zeta^{(i)} x_i - \zeta^{(n+1)} \\ &= \left( \eta^{(1)T} \hat{x}, \dots, \eta^{(s)T} \hat{x} \right)^T \end{aligned} \quad (2.92)$$

and

$$G(x) = \mathbb{P}(\zeta(x) \geq 0).$$

Please note that we distinguish between the random vector  $\zeta$  and  $\zeta(x)$  defined in (2.92).

We assume that  $\zeta$  has a multivariate normal distribution. This implies a multivariate normal distribution for  $\eta$ , as well as for the marginal distributions of  $\zeta^{(j)}$  and  $\eta^{(i)}$ ,  $\forall i, j$ , and for the distribution of  $\zeta(x)$  (see Section 2.2.3 and [328]).

Let  $\mu(x)$  be the expected-value vector and  $\Sigma(x)$  be the covariance matrix of  $\zeta(x)$ . We proceed with computing these moments in terms of the moments of  $\zeta$  and  $\eta$ . To this we introduce some further notation:

- $M$  is the  $(s \times (n+1))$  matrix of expected values of  $(T(\xi), h(\xi))$ ;
- $C^{(i,j)}$  is the  $(s \times s)$  covariance matrix of  $\zeta^{(i)}$  if  $i = j$ , otherwise the cross-covariance matrix of  $\zeta^{(i)}$  and  $\zeta^{(j)}$ ,  $i = 1, \dots, n+1$ ,  $j = 1, \dots, n+1$ ;
- $\tilde{C}^{(i,j)}$  is the  $((n+1) \times (n+1))$  covariance matrix of  $\eta^{(i)}$  if  $i = j$ , otherwise the cross-covariance matrix of  $\eta^{(i)}$  and  $\eta^{(j)}$ ,  $i = 1, \dots, s$ ,  $j = 1, \dots, s$ .

For the expected value of  $\zeta(x)$  we immediately get

$$\mu(x) = \mathbb{E}[\zeta(x)] = \sum_{j=1}^n \mathbb{E}[\zeta^{(j)}] x_j - \mathbb{E}[\zeta^{(n+1)}] = M\hat{x}.$$

For the covariance matrix of  $\zeta(x)$  we obtain two alternative forms corresponding to the column-wise and row-wise representations, respectively. We proceed with the column-wise form. For computing the covariance matrix, we note that  $\zeta(x)$  is defined by an affine linear transformation:

$$\zeta(x) = \begin{pmatrix} x_1 & \dots & x_n & -1 & & \\ & \ddots & & & \ddots & \\ & & x_1 & \dots & x_n & -1 \end{pmatrix} \begin{pmatrix} \zeta^{(1)} \\ \vdots \\ \zeta^{(n)} \\ \zeta^{(n+1)} \end{pmatrix}.$$

A straightforward computation gives for the covariance matrix of  $\zeta(x)$ :

$$\begin{aligned} \Sigma(x) &= \mathbf{Cov}[\zeta(x), \zeta(x)] \\ &= \sum_{i,j=1}^n x_i x_j C^{(i,j)} + \sum_{j=1}^n x_j C^{(n+1,j)} + C^{(n+1,n+1)} \\ &= \sum_{i,j=1}^{n+1} \hat{x}_i \hat{x}_j C^{(i,j)}. \end{aligned} \quad (2.93)$$

In the alternative representation we observe:

$$\zeta(x) = \begin{pmatrix} \hat{x}^T & & \\ & \ddots & \\ & & \hat{x}^T \end{pmatrix} \begin{pmatrix} \eta^{(1)} \\ \vdots \\ \eta^{(s)} \end{pmatrix},$$

which immediately leads to

$$\begin{aligned} \Sigma(x) &= \mathbf{Cov}[\zeta(x), \zeta(x)] \\ &= \begin{pmatrix} \hat{x}^T \bar{C}^{(1,1)} \hat{x} & \dots & \hat{x}^T \bar{C}^{(1,s)} \hat{x} \\ \vdots & \ddots & \vdots \\ \hat{x}^T \bar{C}^{(s,1)} \hat{x} & \dots & \hat{x}^T \bar{C}^{(s,s)} \hat{x} \end{pmatrix}. \end{aligned} \quad (2.94)$$

Next we observe that  $\mathbb{P}(\zeta(x) \geq 0) = \mathbb{P}(-\zeta(x) \leq 0)$ , where  $-\zeta(x)$  is also normally distributed with the same covariance matrix as  $\zeta(x)$  and expected value vector  $-\mu(x)$ . We will consider the case when all covariance matrices  $C^{(i,j)}$  are multiples of a fixed symmetric positive semidefinite matrix. Therefore it is sufficient to prove convexity for one of the sets

$$\begin{aligned} \mathcal{B}(\alpha) &:= \{x \mid \mathbb{P}(\zeta(x) \geq 0) \geq \alpha\}, \\ \mathcal{A}(\alpha) &:= \{x \mid \mathbb{P}(\zeta(x) \leq 0) \geq \alpha\}, \end{aligned}$$

the convexity of the other one follows immediately.

**Theorem 2.12.** *Prékopa [261]. Let us assume that  $\zeta$  has a joint multivariate normal distribution and that*

1. *either there exists an  $((n+1) \times (n+1))$  matrix  $S$  and a symmetric positive semidefinite matrix  $C$ , such that  $C^{(i,j)} = S_{ij}C$  holds,  $\forall i, j$ ,*

2. or there exists an  $(s \times s)$  matrix  $\bar{S}$  and a symmetric positive semidefinite matrix  $\bar{C}$ , such that  $\bar{C}^{(i,j)} = \bar{S}_{ij}\bar{C}$  holds,  $\forall i, j$ .

In both cases, if  $\alpha \geq \frac{1}{2}$  then  $\mathcal{A}(\alpha)$  is a convex set.

*Proof:* We begin with proving the first assertion of the theorem. We will assume that  $S$  is nonsingular; for the general case see Prékopa [266]. For the covariance matrix of  $\zeta(x)$  we have (cf. (2.93))

$$\Sigma(x) = \sum_{i,j=1}^{n+1} \hat{x}_i \hat{x}_j S_{i,j} C = C \cdot \hat{x}^T S \hat{x}. \quad (2.95)$$

In particular, for the variance we get

$$\mathbb{V}\mathbf{ar}[\zeta_i(x)] = C_{ii} \hat{x}^T S \hat{x}, \quad \forall i. \quad (2.96)$$

We may assume that  $C_{ii} > 0$  holds  $\forall i$ .

In fact,  $C_{ii} = 0$  implies that  $\mathbb{V}\mathbf{ar}[\zeta_i(x)] = 0$ ,  $\forall x$ . Consequently, the coefficients and right-hand-side in the  $i^{\text{th}}$  row of the system of random inequalities are a.s. constant. Therefore (see the Remark on page 90) the  $i^{\text{th}}$  inequality can be moved to the set of deterministic constraints in the corresponding optimization problem.

From relation (2.96) immediately follows, that  $S$  is a symmetric positive semidefinite matrix. We have assumed that  $S$  is nonsingular, therefore  $S$  is positive definite.

Another implication of (2.95) is, that the correlation matrix  $R$  of  $\zeta(x)$  does not depend on  $x$ . In fact,  $R_{ij} := \mathbb{C}\mathbf{orr}[\zeta_i(x), \zeta_j(x)] = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}$  holds.

By standardizing  $\zeta(x)$  (see page 99) we get:

$$\mathbb{P}(\zeta(x) \leq 0) = \Phi \left( -\frac{\mu_1(x)}{\sqrt{C_{1,1} \hat{x}^T S \hat{x}}}, \dots, -\frac{\mu_{n+1}(x)}{\sqrt{C_{n+1,n+1} \hat{x}^T S \hat{x}}} ; R \right),$$

where  $\hat{x}^T S \hat{x} > 0$  holds, due to our assumption concerning  $S$  and the fact that  $\hat{x} \neq 0$   $\forall x \in \mathbb{R}^n$ .

Let

$$h^T(x) = \left( -\frac{\mu_1(x)}{\sqrt{C_{1,1}}}, \dots, -\frac{\mu_{n+1}(x)}{\sqrt{C_{n+1,n+1}}} \right),$$

$$\|\hat{z}\|_S = (\hat{z}^T S \hat{z})^{\frac{1}{2}}, \quad \forall \hat{z} \in \mathbb{R}^{n+1},$$

where  $\|\cdot\|_S$  is clearly a norm in  $\mathbb{R}^{n+1}$ . With this notation we have

$$\mathbb{P}(\zeta(x) \leq 0) = \Phi \left( \frac{1}{\|\hat{z}\|_S} h(x); R \right).$$

$\Phi(z; R)$  is a multivariate distribution function, consequently it is monotonically increasing in each of its arguments. This implies that  $\Phi(z_i) \geq \Phi(z; R) \forall i$  holds. Under our assumption  $\alpha \geq \frac{1}{2}$ , we deduce that  $h(x) \geq 0$  holds  $\forall x \in \mathcal{A}(\alpha)$ .

Let  $x \in \mathcal{A}(\alpha)$ ,  $y \in \mathcal{A}(\alpha)$ ,  $\lambda \in (0, 1)$  and let  $\hat{x}$  and  $\hat{y}$  be the corresponding  $(n+1)$ -dimensional vectors with their last coordinate being equal to -1, cf. page 130.

With the notation  $x_\lambda = \lambda x + (1 - \lambda)y$  and  $\hat{x}_\lambda = \lambda \hat{x} + (1 - \lambda)\hat{y}$ , using the triangle inequality for norms we get:

$$\begin{aligned} \mathbb{P}(\zeta(x_\lambda) \leq 0) &= \Phi\left(\frac{1}{\|\hat{x}_\lambda\|_S} h(x_\lambda); R\right) \\ &\geq \Phi\left(\frac{1}{\lambda\|\hat{x}\|_S + (1-\lambda)\|\hat{y}\|_S} (\lambda h(x) + (1-\lambda)h(y)); R\right). \end{aligned} \quad (2.97)$$

We will make use of the following trivial fact: for  $A, B, C, D \in \mathbb{R}$ ,  $C > 0$ , and  $D > 0$  we have

$$\frac{A+B}{C+D} = \kappa \frac{A}{C} + (1-\kappa) \frac{B}{D}$$

with  $\kappa = \frac{C}{C+D}$ ;  $0 < \kappa < 1$ . Applying this componentwise in (2.97) with the setting  $A = \lambda h(x)$ ,  $B = (1-\lambda)h(y)$ ,  $C = \lambda\|\hat{x}\|_S$ , and  $D = (1-\lambda)\|\hat{y}\|_S$ , and utilizing the logconcavity of  $\Phi(z; R)$  we get:

$$\begin{aligned} \mathbb{P}(\zeta(x_\lambda) \leq 0) &\geq \Phi\left(\kappa \frac{1}{\|\hat{x}\|_S} h(x) + (1-\kappa) \frac{1}{\|\hat{y}\|_S} h(y); R\right) \\ &\geq \Phi\left(\frac{1}{\|\hat{x}\|_S} h(x); R\right)^\kappa \Phi\left(\frac{1}{\|\hat{y}\|_S} h(y); R\right)^{1-\kappa} \\ &= \mathbb{P}(\zeta(x) \leq 0)^\kappa \mathbb{P}(\zeta(y) \leq 0)^{1-\kappa} \\ &\geq \alpha^\kappa \alpha^{1-\kappa} = \alpha. \end{aligned} \quad (2.98)$$

The proof of the second assertion runs along analogous lines. For the covariance matrix of  $\zeta(x)$  we now have (see (2.93))

$$\Sigma(x) = \bar{S} \cdot \hat{x}^T \bar{C} \hat{x}. \quad (2.99)$$

For the variance we get

$$\mathbb{V}\text{ar}[\zeta_i(x)] = \bar{S}_{ii} \hat{x}^T \bar{C} \hat{x}, \quad \forall i. \quad (2.100)$$

Arguing similarly as for the first assertion, we conclude that  $\bar{S}_{ii} > 0 \forall i$  may be assumed. If  $\bar{C}$  is positive definite, then the rest of the proof runs analogously to the proof of the first assertion. For the general case see Prékopa [266].  $\square$

Let us remark, that the second assertion of the theorem has originally been proved in [261] under the assumption of the stochastic independence of the rows

of  $(T(\xi), h(\xi))$ ; the general case has been proved by Burkauskas [39].

### 2.2.7 Summary on the convex programming subclasses

SLP models with probability functions are non-convex in general but in the preceding sections we have found important subclasses consisting of convex programming problems. From the practical modeling point of view it is important to know, whether a particular model instance involving probability functions is a convex programming problem. Having namely a convex programming problem there are good chances for finding efficient solution algorithms, or in many cases general-purpose software can be used for solving the problem.

Therefore, for the sake of easy reference, in this section we summarize those model classes which consist of convex programming problems. For further such model classes see Prékopa [266]. If a particular model instance does not belong to any one of these model classes then most probably it is a non-convex optimization problem. This is not certain in general, of course; further research is needed which may lead to the discovery of new convex programming classes of SLP problems with probability functions.

For direct reference we repeat some of the notation and introduce some new one:

$$\begin{aligned} G(x) &= \mathbb{P}_\xi ( T(\xi)x \geq h(\xi) ) \\ \hat{G}(x) &= \mathbb{P}_\xi ( T(\xi)x \leq h(\xi) ) \\ \mathcal{B}(\alpha) &= \{ x \mid G(x) \geq \alpha \} \\ \hat{\mathcal{B}}(\alpha) &= \{ x \mid \hat{G}(x) \geq \alpha \} \\ \mathcal{D}(\beta) &= \{ x \mid G(x) \leq \beta \} \\ \hat{\mathcal{D}}(\beta) &= \{ x \mid \hat{G}(x) \leq \beta \}, \end{aligned}$$

where  $T(\xi)$  denotes a random  $s \times n$  matrix,  $h(\xi) \in \mathbb{R}^s$  stands for a random vector. The components of the  $n$ -dimensional random vector  $t_i(\xi)$  are the elements of the  $i^{\text{th}}$  row of  $T(\xi)$ ,  $\forall i$  and  $T_j(\xi)$  stands for the  $j^{\text{th}}$  column of  $T(\xi)$ ,  $\forall j$ . If  $s = 1$  (separate probability function) holds, we use the notation  $t(\xi) = T(\xi)$ ;  $\mu = \mathbb{E}[t(\xi)]$ ,  $\mu_{t+1} = \mathbb{E}[h(\xi)]$ .

- A. General cases: convex models are identified by choosing specific probability levels. If  $\alpha = 1$  or  $\alpha = 0$  or  $\beta = 1$  then  $\mathcal{B}(\alpha)$ ,  $\hat{\mathcal{B}}(\alpha)$ ,  $\mathcal{D}(\beta)$ , and  $\hat{\mathcal{D}}(\beta)$  are all convex sets. (Proposition 2.2. on page 92 and the discussion on page 92).
- B.  $\xi$  has a finite discrete distribution: convex models are identified by choosing specific probability levels. If  $\alpha$  is high enough (as precisely formulated in the assumptions of Proposition 2.3. on page 95 and Proposition 2.4. on page 95) then  $\mathcal{B}(\alpha)$  and  $\hat{\mathcal{B}}(\alpha)$  are convex. In general, however,  $\mathcal{B}(\alpha)$ ,  $\hat{\mathcal{B}}(\alpha)$ ,  $\mathcal{D}(\beta)$ , and  $\hat{\mathcal{D}}(\beta)$  are non-convex sets. Equivalent linear mixed-integer programming reformulations are available, see (2.44) on page 94.



C. Separate probability functions,  $s = 1$ : convex cases are identified by choosing specific probability distributions and probability levels.

1. If only the right-hand-side is stochastic then  $\mathcal{B}(\alpha)$ ,  $\widehat{\mathcal{B}}(\alpha)$ ,  $\mathcal{D}(\beta)$ , and  $\widehat{\mathcal{D}}(\beta)$  are half-spaces, determined by linear inequalities (Section 2.2.3) although for  $\mathcal{D}(\beta)$ , and  $\widehat{\mathcal{D}}(\beta)$  some care is needed if  $\xi$  does not have a continuous distribution (page 98). (2.91) can be formulated as a deterministic linear program if  $\mathcal{B}$  is bounded, otherwise some caution is needed, see (2.49) on page 98. These results hold for arbitrary values  $0 < \alpha < 1$  and  $0 < \beta < 1$ .
2. If  $(t(\xi), h(\xi))$  has a multivariate normal distribution and  $\alpha \geq \frac{1}{2}$  and  $\beta \leq \frac{1}{2}$  hold, then  $\mathcal{B}(\alpha)$ ,  $\widehat{\mathcal{B}}(\alpha)$ ,  $\mathcal{D}(\beta)$ , and  $\widehat{\mathcal{D}}(\beta)$  are convex sets, determined by convex nonlinear constraints, see Section 2.2.3, (2.60), and (2.61) on page 103. Some care is needed concerning  $\mathcal{D}(\beta)$  and  $\widehat{\mathcal{D}}(\beta)$ , see (2.61) on page 103, in the case when the distribution is degenerate, see page 103. If the distribution is non-degenerate then  $G(x)$  is pseudo-concave on  $\mathcal{B}$ , if  $\mu^T x \geq \mu_{n+1}$  holds for all  $x \in \mathcal{B}$ . It is pseudo-convex on  $\mathcal{B}$ , if  $\mu^T x \leq \mu_{n+1}$  holds for all  $x \in \mathcal{B}$ . Similar assertions hold for  $\widehat{G}(x)$  with exchanged roles of the inequalities for the expected values (Proposition 2.36.) on page 104. Thus (2.91) is a convex programming problem if for all  $x \in \mathcal{B}$ ,  $\mu^T x \geq \mu_{n+1}$  holds. The corresponding minimization problem is a convex programming problem provided that for all  $x \in \mathcal{B}$  the reverse strict inequalities  $\mu^T x \leq \mu_{n+1}$  hold.
3. If  $(t(\xi), h(\xi))$  has a multivariate Cauchy distribution, similar remarks apply as in the normally distributed case, see Section 2.2.3. This section outlines also a technique for carrying out the analysis for distributions belonging to the class of stable distributions.

D. Stochastically independent random variables,  $s > 1$ : convex cases are identified by choosing specific probability distributions. If only  $h(\xi)$  is stochastic,  $(h_1(\xi), \dots, h_s(\xi))$  are stochastically independent, and each  $h_i(\xi)$  has a log-normal distribution function, then  $G(x)$  is a logconcave function and  $\mathcal{B}(\alpha)$  is convex (Section 2.2.4).

E. Only the right-hand-side is stochastic: convex cases are identified by choosing specific probability distributions. In the case of  $s = 1$  this has been discussed above in item B.1 and under the assumption of stochastic independence the discussion can be found under item D. In the general case  $G(x)$  and  $\widehat{G}(x)$  are log-concave for the following multivariate distributions: uniform (page 122), non-degenerate normal (page 122), Dirichlet (page 123), Wishart (page 123), log-normal (page 126), and gamma (page 127). The probability functions  $G$  and  $\widehat{G}(x)$  are quasi-concave for the multivariate t-distribution (page 124). Consequently,  $\mathcal{B}(\alpha)$  and  $\widehat{\mathcal{B}}(\alpha)$  are convex. Having  $G(x)$  or  $\widehat{G}(x)$  in the objective function, (2.91) is a convex programming problem for the logconcave distributions listed above. Regarding the case with reverse inequality constraints and the same distributions,  $\mathcal{D}(\beta)$ , and  $\widehat{\mathcal{D}}(\beta)$  are non-convex sets in general and the minimization variant of (2.91) is a non-convex optimization problem.

F. Random technology matrix: for the case  $s = 1$  the discussion can be found under items C.2 and C.3. For  $s > 1$ ,  $\mathcal{B}(\alpha)$  and  $\widehat{\mathcal{B}}(\alpha)$  are convex under the following assumptions:  $(T_1(\xi), \dots, T_n(\xi), h(\xi))$  have a joint multivariate normal distribution and the covariance matrices of the columns as well as the cross-covariance matrices are constant multiples of a fixed covariance matrix, then  $\mathcal{B}(\alpha)$  and  $\widehat{\mathcal{B}}(\alpha)$  are convex sets. This holds also under the analogous assumption concerning the rows. For both facts see Proposition 2.12. on page 131.

## Exercises

**2.4.** Show the following assertion concerning probabilities, which has been utilized in the proof of Theorem 2.2. on page 92:

$$\mathbb{P}(A) = \mathbb{P}(B) = 1 \Rightarrow \mathbb{P}(A \cap B) = 1.$$

**2.5.** Consider the following chance-constrained problem:

$$\left. \begin{array}{ll} \min & 2x_1 + x_2 \\ \text{s.t.} & x_1 \geq 1 \\ & \mathbb{P}(x_1 + x_2 \geq \xi) \geq 0.9, \end{array} \right\}$$

with  $\xi \in \mathcal{U}(1, 2)$ , meaning that the random variable  $\xi$  is uniformly distributed over the interval  $[1, 2]$ .

- (a) Formulate the equivalent LP problem.
- (b) Solve this LP graphically and solve the original problem by employing SLP-IOR; compare the results.

**2.6.** In the proof of Proposition 2.36. on page 104 we have utilized the fact that  $h(x) := \|D^T x - d\|$  is a convex function. Prove that this holds for any norm  $\|\cdot\|$ .

**2.7.** Consider the following pair of chance-constrained problems:

$$\left. \begin{array}{ll} \min & 2x_1 + x_2 \\ \text{s.t.} & x_1 + x_2 \leq 8 \\ & \mathbb{P}(x_1 + x_2 \geq \xi_1) \geq 0.95 \\ & \mathbb{P}(x_1 \geq \xi_2) \geq 0.95 \\ & x_1, x_2 \geq 0 \end{array} \right\} (S) \quad \left. \begin{array}{ll} \min & 2x_1 + x_2 \\ \text{s.t.} & x_1 + x_2 \leq 8 \\ & \mathbb{P} \left( \begin{array}{l} x_1 + x_2 \geq \xi_1 \\ x_1 \geq \xi_2 \end{array} \right) \geq 0.95 \\ & x_1, x_2 \geq 0, \end{array} \right\} (J)$$

where the probability distribution of  $\xi = (\xi_1, \xi_2)^T$  is a normal distribution with parameters  $\mathbb{E}[\xi] = (2, 1)^T$ , standard deviations  $\sigma[\xi] = (0.5, 0.5)^T$  and correlation

$\sigma_{1,2}(\xi_1, \xi_2) = 0.2$ . Problems  $(S)$  and  $(J)$  are formulated on the basis of the same data-set,  $(S)$  with separate chance constraints and  $(J)$  with a joint chance constraint. In  $(S)$  the marginal distributions of  $(\xi_1, \xi_2)$  are chosen for the probability distributions of  $\xi_1$  and  $\xi_2$ . Notice that for both of the separate constraints the same probability level is prescribed as for the joint constraint.

- (a) Solve both problems by utilizing SLP–IOR and compare the optimal objective values  $z_S^*$  and  $z_J^*$ .
- (b) It will turn out that  $z_S^* < z_J^*$  holds. Show that this is not just by chance: if formulating two chance-constrained problems on the same data-set in the above way, the optimal (minimal) objective value for the problem with separate constraints never exceeds the optimal objective value of the optimization problem with joint constraints.

## 2.3 Quantile functions, Value at Risk

One way for including simultaneously the loss size and the probability of loss into an SLP model leads via quantiles. Recall that for a random variable  $\vartheta$  with distribution function  $F_\vartheta$  and for  $0 < \alpha < 1$ ,  $z \in \mathbb{R}$  is an  $\alpha$ -quantile, if both inequalities

$$\mathbb{P}(\vartheta \leq z) \geq \alpha \quad \text{and} \quad \mathbb{P}(\vartheta \geq z) \geq 1 - \alpha$$

hold. The set of  $\alpha$ -quantiles is a non-empty closed interval for  $0 < \alpha < 1$ , see, for instance, Cramér [47]. We assume that  $0 < \alpha < 1$  holds and assign the following quality measure to random variables:

$$\rho_{\text{VaR}}^\alpha(\vartheta) := v(\vartheta, \alpha) := \min\{z \mid F_\vartheta(z) \geq \alpha\}, \quad \vartheta \in \mathcal{L}_1^0, \quad (2.101)$$

defined on the set of all random variables over  $\Omega$ . According to this definition, for a given  $\alpha$ ,  $v(\vartheta, \alpha)$  is the left endpoint of the closed interval of  $\alpha$ -quantiles of  $\vartheta$ .

Similarly as in Section 2.2.3 on separate probability functions, for the sake of simplicity of notation, we consider the random variable

$$\zeta(x, \eta, \xi) := \eta^T x - \xi. \quad (2.102)$$

We interpret positive values of  $\zeta(x, \eta, \xi)$  as loss and negative values as gain. The evaluation function corresponding to the risk measure (2.101) will be the following:

$$v(x, \alpha) := \min\{z \mid \Psi(x, z) \geq \alpha\}, \quad (2.103)$$

where  $\Psi(x, \cdot)$  denotes the probability distribution function of  $\zeta(x, \eta, \xi)$ . We will call  $v(x, \alpha)$  a *quantile function*.  $\alpha$  will typically have a large value, for instance,  $\alpha = 0.95$ . The interpretation of  $v(x, \alpha)$  is in this case a minimal loss level, corresponding to the decision vector  $x$ , with the following property: the probability of the event that the loss will not exceed  $v(x, \alpha)$  is at least  $\alpha$ . In financial applications  $v(x, \alpha)$  is

called Value at Risk (VaR), see Elton et al. [85], and the references therein. We will adopt this terminology for our more general setting.

We consider minimizing the sum of a linear function and VaR, under linear constraints:

$$\left. \begin{array}{ll} \min & c^T x + v(x, \alpha) \\ \text{s.t.} & x \in \mathcal{B}. \end{array} \right\} \quad (2.104)$$

By using the definition of  $v(x, \alpha)$  and introducing an additional variable  $z$ , the following equivalent formulation results:

$$\left. \begin{array}{ll} \min & c^T x + z \\ \text{s.t.} & \Psi(x, z) \geq \alpha \\ & x \in \mathcal{B}. \end{array} \right\} \quad (2.105)$$

The equivalence with (2.104) is immediate by noting that for each fixed  $x \in \mathcal{B}$  in (2.105), it is sufficient to take into account the minimal  $z$  in the constraint, this minimal  $z$  is however  $v(x, \alpha)$ . Substituting the definition of  $\Psi$  finally leads to the formulation

$$\left. \begin{array}{ll} \min & c^T x + z \\ \text{s.t.} & \mathbb{P}(\eta^T x - \xi \leq z) \geq \alpha \\ & x \in \mathcal{B}. \end{array} \right\} \quad (2.106)$$

This model clearly belongs to the class of SLP models with separate probability functions, see (2.46) with  $s = 1$ , in Section 2.2.3. The probability function in the model above is a special case of the general form with the “technology vector” containing a deterministic component

$$\hat{G}(x, z) = \mathbb{P} \left( (\eta^T, -1) \begin{pmatrix} x \\ z \end{pmatrix} - \xi \leq 0 \right).$$

It is an interesting fact, that the first SLP model for minimizing VaR has been formulated by Kataoka [178] in the form (2.106) already in 1963.

Being a special case of SLP models with separate probability functions, the whole machinery developed in Section 2.2.3 applies. We will illustrate this by discussing the case of the multivariate normal distribution. Let

$$\zeta(x, z, \eta, \xi) := \eta^T x - \xi - z$$

and assume that  $(\eta, \xi)$  has a multivariate normal distribution (see page 100). For a fixed  $(x, z)$ , the  $z$ -term can be interpreted as merely modifying the expected value of  $\xi$ , therefore for  $\hat{G}(x, z)$  the explicit form (2.59) on page 102 applies with  $\mu_{n+1}$  replaced by  $\mu_{n+1} + z$ . Consequently, see (2.60) on page 102, (2.106) can be written as

$$\left. \begin{array}{ll} \min & c^T x + z \\ \text{s.t.} & \Phi^{-1}(\alpha) \|D^T x - d\| + \mu^T x - z \leq \mu_{n+1} \\ & x \in \mathcal{B}. \end{array} \right\} \quad (2.107)$$

At the optimal solution the nonlinear constraint is clearly active. This observation leads, by eliminating  $z$ , to the following linearly constrained alternative formulation:

$$\left. \begin{array}{ll} \min_x & c^T x + \Phi^{-1}(\alpha) \|D^T x - d\| + \mu^T x - \mu_{n+1} \\ \text{s.t.} & x \in \mathcal{B}. \end{array} \right\} \quad (2.108)$$

Assuming that  $\alpha \geq \frac{1}{2}$  holds, due to the convexity of the Euclidean norm both models (2.107) and (2.108) are convex programming problems.

Except of those cases, discussed in Section 2.2.3, which can be formulated as convex programming problems, the model (2.106) is in general a non-convex optimization problem.

Turning now our attention to SLP problems with VaR-constraints, we consider problems of the following form:

$$\left. \begin{array}{ll} \min_x & c^T x \\ \text{s.t.} & \min_z \{z \mid \Psi(x, z) \geq \alpha\} \leq \kappa \\ & x \in \mathcal{B}. \end{array} \right\} \quad (2.109)$$

Observe that the minimum in the minimization problem involved in the first constraint is attained. Therefore, for a fixed  $x$  this constraint holds, if and only if there exists a  $z \in \mathbb{R}$  such that it holds for that  $z$ . Thus the optimization problem (2.109) can be equivalently formulated as follows:

$$\left. \begin{array}{ll} \min_{x, z} & c^T x \\ \text{s.t.} & \Psi(x, z) \geq \alpha \\ & z \leq \kappa \\ & x \in \mathcal{B}. \end{array} \right\} \quad (2.110)$$

Finally, substituting the definition of  $\Psi$  results in

$$\left. \begin{array}{ll} \min_{x, z} & c^T x \\ \text{s.t.} & \mathbb{P}(\eta^T x - z \leq \xi) \geq \alpha \\ & z \leq \kappa \\ & x \in \mathcal{B}. \end{array} \right\} \quad (2.111)$$

Thus, also in this case, we have obtained an equivalent problem which belongs to the class of SLP problems with separate probability functions, see Section 2.2.3. Therefore, analogous comments and formulations apply, as for the SLP problem in which VaR is minimized, see (2.106).

For further stochastic programming problems based on quantile functions see Kibzun and Kan [182].

## 2.4 Models based on expectation

The simplest way of including expectations into an SLP model is based on choosing the quality measure

$$\rho_E(\vartheta) := \mathbb{E}[\vartheta], \quad \vartheta \in \mathcal{L}_1^1,$$

defined on the linear space of random variables with finite expected value. We consider the random variable

$$\zeta(x, \xi) = t^T(\xi)x - h(\xi),$$

where  $t(\xi)$  is an  $n$ -dimensional random vector and  $h(\xi)$  is a random variable. Under the assumption that the expected value of  $(T(\xi), h(\xi))$  exists, we obtain the following deterministic linear-affine evaluation function for  $x$ :

$$\mathbb{E}[\zeta(x, \xi)] = \bar{t}^T x - \bar{h}$$

with  $\bar{t} = \mathbb{E}[t(\xi)]$  and  $\bar{h} = \mathbb{E}[h(\xi)]$ . In the case when  $\zeta(x, \xi)$  is a random vector, this holds componentwise. Consequently, the prototype models (2.7) and (2.9) become linear programming problems. These LP's are called *expected value problems*, corresponding to the SLP problem.

On the one hand, having an equivalent linear programming problem is an attractive feature from the numerical point of view. On the other hand, however, replacing the probability distribution by a one-point distribution leads to a very crude approximation of the original distribution in general. In some modeling situations it may happen that the solution  $\bar{x}$  of the expected value problem also solves a corresponding SLP problem. However, this is usually an indication of a modeling or data error: the corresponding SLP model is not “truly stochastic”. Unfortunately, the expected value problem is frequently used by modelers as a substitute for the SLP problem, without further considerations. When doing this, extreme care is needed, since the solution obtained this way may turn out to be quite risky when evaluated by an alternative evaluation function. Taking the expected value problem should by no means be used as the single way representing  $\zeta(x, \xi)$  in the model. Accompanied with other constraints or objective functions, based on alternative quality measures, utilizing  $\rho_E(\vartheta)$  may lead to important and meaningful model formulations. As an example we refer to the portfolio optimization model of Markowitz [217] which has been applied with tremendous success in finance.

The picture radically changes if the expectation is taken separately for the positive- or negative part of  $\zeta(x, \xi)$ , or if conditional expectations are utilized. In this section we will discuss several important model classes based on these ideas.

We shall need some basic facts from probability theory concerning expectations. Let  $\vartheta$  be a random variable and assume that  $\mathbb{E}[\vartheta]$  exists. Recall from probability theory, that this assumption means the finiteness of the integral  $\int_{-\infty}^{\infty} |t| dF_{\vartheta}(t)$ , where  $F_{\vartheta}$  denotes the probability distribution of  $\vartheta$ .

The following well-known integral representations will be used in this section, for which, for the sake of completeness, we also present a proof. Introducing the notation  $u^+ := \max\{0, u\}$  and  $u^- := \max\{0, -u\}$  for all  $u \in \mathbb{R}$ , we have

**Proposition 2.38.** *Assume that  $\mathbb{E}[\vartheta]$  exists. Then for all  $z \in \mathbb{R}$  both  $\mathbb{E}[(\vartheta - z)^+]$  and  $\mathbb{E}[(\vartheta - z)^-]$  exist and we have:*

$$\begin{aligned}\mathbb{E}[(\vartheta - z)^+] &= \int_z^\infty (1 - F_\vartheta(t)) dt \\ \mathbb{E}[(\vartheta - z)^-] &= \int_{-\infty}^z F_\vartheta(t) dt.\end{aligned}\tag{2.112}$$

*Proof:* The existence of  $\mathbb{E}[\vartheta]$  obviously implies the existence of the expected values on the left-hand-side in (2.112). Using integration by parts we get for  $z < y$

$$\begin{aligned}\int_z^y (t - z) dF_\vartheta(t) &= (t - z)F_\vartheta(t) \Big|_z^y - \int_z^y F_\vartheta(t) dt \\ &= -(y - z)(1 - F_\vartheta(y)) + \int_z^y (1 - F_\vartheta(t)) dt\end{aligned}$$

and consequently

$$\begin{aligned}\mathbb{E}[(\vartheta - z)^+] &= \int_z^\infty (t - z) dF_\vartheta(t) \\ &= \lim_{y \rightarrow \infty} \int_z^y (t - z) dF_\vartheta(t) = \int_z^\infty (1 - F_\vartheta(t)) dt,\end{aligned}$$

where we have used the fact that the existence of the expected value of  $\vartheta$  implies that  $\lim_{y \rightarrow \infty} y(1 - F_\vartheta(y)) = 0$  holds. For the second relation we get similarly via integration by parts:

$$\begin{aligned}\mathbb{E}[(\vartheta - z)^-] &= \int_{-\infty}^z (z - t) dF_\vartheta(t) \\ &= (z - t)F_\vartheta(t) \Big|_{-\infty}^z + \int_{-\infty}^z F_\vartheta(t) dt = \int_{-\infty}^z F_\vartheta(t) dt,\end{aligned}$$

where we used that  $\lim_{x \rightarrow -\infty} xF_\vartheta(x) = 0$  holds, due to the existence of the expected value of  $\vartheta$ . □

### 2.4.1 Integrated chance constraints

Similarly as in Section 2.2 concerning probability functions, also in this section we will distinguish two cases: first we discuss the case when  $\zeta(x, \xi) := T(\xi)x - h(\xi)$  is a random variable ( $s = 1$  holds, see (2.1) on page 71). Afterwards we consider the general case when  $\zeta(x, \xi) := T(\xi)x - h(\xi)$  is a random vector, that means,  $s \geq 1$  holds. We will assume throughout that the expected values of  $T(\xi)$  and  $h(\xi)$  exist.

#### Separate integrated probability functions

We consider the random variable

$$\zeta(x, \xi) := t(\xi)^T x - h(\xi),$$

where  $t(\xi)$  is an  $n$ -dimensional random vector and  $h(\xi)$  is a random variable. Depending on whether positive or negative values of  $\zeta(x, \xi)$  are considered as losses, the loss as a random variable can be written as

$$\zeta^+(x, \xi) := [t(\xi)^T x - h(\xi)]^+$$

or

$$\zeta^-(x, \xi) := [t(\xi)^T x - h(\xi)]^-,$$

respectively. Here we have made use of the notation  $z^+ = \max\{0, z\}$  and  $z^- = \max\{0, -z\}$ ,  $z \in \mathbb{R}$ .  $z^+$  will be called the positive part and  $z^-$  the negative part of the real number  $z$ .

For being in accordance with the literature, let us assume that losses are modeled as negative values of  $\zeta(x, \xi)$ . Using the notation above, the probability constraint corresponding to the random linear inequality  $\zeta(x, \xi) \geq 0$  can obviously be written in expectation terms (see (2.33) on page 90) as follows

$$\mathbb{P}_\xi(\zeta(x, \xi) \geq 0) \geq \alpha \iff \mathbb{E}_\xi[\chi(\zeta^-(x, \xi))] \leq 1 - \alpha \quad (2.113)$$

with the indicator function

$$\chi(z) := \begin{cases} 0 & \text{if } z \leq 0, \\ 1 & \text{if } z > 0. \end{cases}$$

In the second inequality in (2.113) the application of the function  $\chi$  results in assigning the constant value 1 to the loss irrespectively of its size. This can heuristically be viewed as the source of the generally non-convex behavior of probability functions, see the nice examples in Klein Haneveld [188] and Klein Haneveld and Van der Vlerk [191]. This observation leads to integrated chance constraints by dropping  $\chi$  in (2.113) and by prescribing an upper bound for  $\mathbb{E}_\xi[\zeta^-(x, \xi)]$ . More specifically, we choose two risk measures for random variables as



$$\begin{aligned}\rho_{\text{sic}}^+(\vartheta) &:= \mathbb{E}[\vartheta^+], \\ \rho_{\text{sic}}^-(\vartheta) &:= \mathbb{E}[\vartheta^-], \quad \vartheta \in \mathcal{L}_1^1,\end{aligned}\tag{2.114}$$

defined on the linear space of random variables with finite expected value. The corresponding evaluation functions  $K$  and  $H$  will be

$$\begin{aligned}K(x) &:= \rho_{\text{sic}}^+(\zeta(x, \xi)) = \mathbb{E}_\xi[\zeta^+(x, \xi)] \quad \text{and} \\ H(x) &:= \rho_{\text{sic}}^-(\zeta(x, \xi)) = \mathbb{E}_\xi[\zeta^-(x, \xi)],\end{aligned}$$

respectively. The functions  $K(x)$  and  $H(x)$  will be called *separate integrated probability functions*. Assuming, for instance, that losses correspond to negative values of  $\zeta(x, \xi)$ , a *separate integrated chance constraint* has the form

$$\mathbb{E}_\xi[(\zeta(x, \xi))^-] \leq \gamma, \tag{2.115}$$

where  $\gamma$  is a prescribed maximal tolerable expected loss. The following relation provides an explanation of the term “integrated”: due to Proposition 2.38. on page 141 we have

$$\mathbb{E}_\xi[(\zeta(x, \xi))^-] = \int_{-\infty}^0 \mathbb{P}(\zeta(x, \xi) \leq z) dz = \int_{-\infty}^0 \mathbb{P}(\zeta(x, \xi) < z) dz.$$

The second equality holds because the set of jump-points of the distribution function  $\Psi(x, \cdot) := \mathbb{P}(\zeta(x, \xi) \leq z)$  is countable and therefore it has (Lebesgue) measure 0.

Let us define the positive- and negative-part functions  $\varphi^+$  and  $\varphi^-$  according to  $\varphi^+(z) := z^+$  and  $\varphi^-(z) := z^-$  for  $z \in \mathbb{R}$ , respectively. Both of these functions are obviously convex. From the optimization point of view the most attractive property of separate integrated probability functions is formulated in the subsequent proposition:

**Proposition 2.39.** *Both  $H(x)$  and  $K(x)$ , and consequently  $\mathbb{E}_\xi[|\zeta(x, \xi)|] = H(x) + K(x)$  are convex functions on  $\mathbb{R}^n$ .*

*Proof:* The assertion follows easily from the convexity of the functions  $\varphi^+(\cdot)$  and  $\varphi^-(\cdot)$ . We prove the assertion for  $K(x)$ ; the proof for  $H(x)$  is analogous. We have  $K(x) = \mathbb{E}_\xi[\varphi^+(\zeta(x, \xi))]$ . Because  $\zeta(x, \xi)$  is linear in  $x$  and  $\varphi^+$  is a convex function,  $\varphi^+(\zeta(x, \xi))$  is convex for each fixed  $\xi$ . Taking the expected value preserves convexity. For a formal proof let  $x, y \in \mathbb{R}^n$  and  $0 \leq \lambda \leq 1$ . We have

$$\begin{aligned}K(\lambda x + (1 - \lambda)y) &= \mathbb{E}_\xi[\varphi^+(\zeta(\lambda x + (1 - \lambda)y, \xi))] \\ &= \mathbb{E}_\xi[\varphi^+(\lambda \zeta(x, \xi) + (1 - \lambda)\zeta(y, \xi))] \\ &\leq \mathbb{E}_\xi[\lambda \varphi^+(\zeta(x, \xi)) + (1 - \lambda)\varphi^+(\zeta(y, \xi))] \\ &= \lambda K(x) + (1 - \lambda)K(y).\end{aligned}$$

□

This result implies that  $K(x)$  and  $H(x)$  are convex, in particular also for finite discrete distributions. This is in sharp contrast with probability functions, where (generalized) concavity holds only under various assumptions, excluding finite discrete distributions in general.

With  $\zeta^+(x, \xi)$  representing losses, the following prototype models will be considered:

$$\left. \begin{array}{ll} \min & c^T x \\ \text{s.t.} & \mathbb{E}_\xi[\zeta^+(x, \xi)] \leq \gamma \\ & x \in \mathcal{B} \end{array} \right\} \quad (2.116)$$

and

$$\left. \begin{array}{ll} \min & c^T x + \mathbb{E}_\xi[\zeta^+(x, \xi)] \\ \text{s.t.} & x \in \mathcal{B}, \end{array} \right\} \quad (2.117)$$

where  $\gamma > 0$  is a prescribed maximally tolerable loss level. Due to Proposition 2.39., both problems are convex programming problems. Convex functions being continuous (see, for instance, Rockafellar [281]), the feasible set of (2.116) is obviously closed.

Note that there is no way of building convex programming models of the above type with reversed inequality constraints in (2.116) or with maximization in (2.117) which are based on separate integrated probability functions. Because both  $K(x)$  and  $H(x)$  are convex, it is immaterial whether the loss is represented by  $\zeta^+(x, \xi)$  or by  $\zeta^-(x, \xi)$ .

Next we assume that  $\xi$  has a finite discrete distribution with  $N$  realizations and corresponding probabilities given in the tableau

$$\begin{pmatrix} p_1 & \dots & p_N \\ \hat{\xi}^1 & \dots & \hat{\xi}^N \end{pmatrix} \quad (2.118)$$

with  $p_i > 0 \forall i$  and  $\sum_{i=1}^N p_i = 1$ . We introduce the notation  $T^k = T(\hat{\xi}^k)$ ,  $h^k = h(\hat{\xi}^k)$ ,  $k = 1, \dots, N$ , and  $\mathcal{N} = \{1, \dots, N\}$ . The  $i^{\text{th}}$  row of  $T^k$  will be denoted by  $t_i^k$  and if  $s = 1$  then the single row of  $T^k$  will be denoted by  $t^k$ . For notational convenience, both  $t_i^k$  and  $t^k$  will be considered as row-vectors ( $(1 \times n)$  matrices).

Problems (2.116) and (2.117) can in this case be formulated as follows:

$$\left. \begin{array}{ll} \min & c^T x \\ \text{s.t.} & \sum_{k=1}^N p_k (t^k x - h^k)^+ \leq \gamma \\ & x \in \mathcal{B} \end{array} \right\} \quad (2.119)$$

and

$$\left. \begin{array}{l} \min c^T x + \sum_{k=1}^N p_k (t^k x - h^k)^+ \\ \text{s.t. } x \in \mathcal{B}. \end{array} \right\} \quad (2.120)$$

These nonlinear programming problems can be equivalently formulated as linear programming problems by introducing the auxiliary variables  $y^k, k = 1, \dots, N$ :

$$\left. \begin{array}{l} \min c^T x \\ \text{s.t. } \sum_{k=1}^N p_k y^k \leq \gamma \\ t^k x - y^k \leq h^k, k = 1, \dots, N \\ y^k \geq 0, k = 1, \dots, N \\ x \in \mathcal{B} \end{array} \right\} \quad (2.121)$$

and

$$\left. \begin{array}{l} \min c^T x + \sum_{k=1}^N p_k y^k \\ \text{s.t. } t^k x - y^k \leq h^k, k = 1, \dots, N \\ y^k \geq 0, k = 1, \dots, N \\ x \in \mathcal{B}. \end{array} \right\} \quad (2.122)$$

The equivalence of (2.119) and (2.121) as well as the equivalence of (2.120) and (2.122) follows easily from the following fact: if  $\bar{x}, \bar{y}^k, k = 1, \dots, N$  is a feasible solution of either (2.121) or (2.122), then the following inequality holds:

$$\sum_{k=1}^N p_k (t^k \bar{x} - h^k)^+ \leq \sum_{k=1}^N p_k \bar{y}^k.$$

Let  $\mathcal{S}(\gamma) = \{x \mid \mathbb{E}_\xi[\zeta^+(x, \xi)] \leq \gamma\}$  be the set of feasible solutions corresponding to the integrated chance constraint. The following representation holds, which plays an important role in the dual decomposition algorithm (see Section 4.4.3).

**Theorem 2.13.** *Klein Haneveld and Van der Vlerk [191]. For  $\gamma \geq 0$ ,  $\mathcal{S}(\gamma)$  is a polyhedral set. In fact the following representation holds:*

$$\{x \mid K(x) \leq \gamma\} = \bigcap_{\mathcal{H} \subset \mathcal{N}} \{x \mid \sum_{k \in \mathcal{H}} p_k (t^k x - h^k) \leq \gamma\}. \quad (2.123)$$

*Proof:* We have

$$\begin{aligned}
K(x) &= \mathbb{E}_\xi[\zeta^+(x, \xi)] = \sum_{k=1}^N p_k \zeta^+(x, \widehat{\xi}^k) \\
&= \sum_{k: \zeta(x, \widehat{\xi}^k) > 0} p_k \zeta(x, \widehat{\xi}^k) \\
&= \max_{\mathcal{H} \subset \mathcal{N}} \sum_{k \in \mathcal{H}} p_k \zeta(x, \widehat{\xi}^k).
\end{aligned}$$

Using this representation we get

$$\mathcal{S}(\gamma) = \{x \mid K(x) \leq \gamma\} = \bigcap_{\mathcal{H} \subset \mathcal{N}} \{x \mid \sum_{k \in \mathcal{H}} p_k \zeta(x, \widehat{\xi}^k) \leq \gamma\}$$

from which the result immediately follows.  $\square$

In (2.123), for  $\mathcal{H} = \emptyset \subset \mathcal{N}$  the sum over the empty index set is interpreted as having the value 0 thus the corresponding inequality holds for any  $x$ . Consequently,  $\mathcal{S}(\gamma)$  is represented by a system of  $2^N - 1$  proper linear inequalities. Models (2.116) and (2.117) deliver identical solutions for random variables for which  $\zeta^+(x, \xi)$  is the same almost surely. This is not the case with the following variant of integrated chance constraints:

$$\mathbb{E}_\xi[\zeta^+(x, \xi)] \leq \alpha \mathbb{E}_\xi[|\zeta(x, \xi)|] \quad (2.124)$$

with  $\alpha$  being prescribed. Because  $K(x) \geq 0$  and  $K(x) \leq \mathbb{E}_\xi[|\zeta(x, \xi)|]$  obviously hold for all  $x$ , it is sufficient to consider  $\alpha$ -values with  $\alpha \in [0, 1]$ . Using the relations  $z = z^+ - z^-$  and  $|z| = z^+ + z^-$ ,  $z \in \mathbb{R}$ , the above inequality can be equivalently written as

$$(1 - 2\alpha)\mathbb{E}_\xi[\zeta^+(x, \xi)] + \alpha\mathbb{E}_\xi[\zeta(x, \xi)] \leq 0 \quad (2.125)$$

or as

$$(1 - 2\alpha)\mathbb{E}_\xi[\zeta^+(x, \xi)] + \alpha(\bar{t}x - \bar{h}) \leq 0 \quad (2.126)$$

with  $\bar{t} = \mathbb{E}_\xi[t(\xi)]$  and  $\bar{h} = \mathbb{E}_\xi[h(\xi)]$ .

This motivates the choice of the following quality measure for evaluating random variables in our framework for constructing SLP models:

$$\rho_{\text{sic}}^\alpha(\vartheta) := (1 - \alpha)\mathbb{E}[\vartheta^+] - \alpha\mathbb{E}[\vartheta^-] = \alpha\mathbb{E}[\vartheta] + (1 - 2\alpha)\mathbb{E}[\vartheta^+], \quad \vartheta \in \mathcal{L}_1^1.$$

We obtain the evaluation function as usual by substituting  $\vartheta = \zeta(x, \xi)$ :

$$K_\alpha(x) := \alpha(\bar{t}x - \bar{h}) + (1 - 2\alpha)\mathbb{E}_\xi[\zeta^+(x, \xi)].$$

Proposition 2.39. implies that  $K_\alpha(x)$  is convex for  $\alpha < \frac{1}{2}$  and it is concave for  $\alpha > \frac{1}{2}$ . For  $\alpha = \frac{1}{2}$  the function is clearly linear-affine.

Choosing  $\alpha$  such that  $\alpha \in [0, \frac{1}{2}]$  holds, the parameter  $\alpha$  will be interpreted as a risk-aversion parameter. Decreasing  $\alpha$  means increasing risk-aversion. The prototype models will have the form

$$\left. \begin{array}{ll} \min & c^T x \\ \text{s.t.} & \alpha(\bar{t}x - \bar{h}) + (1 - 2\alpha)\mathbb{E}_\xi[\zeta^+(x, \xi)] \leq 0 \\ & x \in \mathcal{B} \end{array} \right\} \quad (2.127)$$

and

$$\left. \begin{array}{ll} \min & c^T x + \alpha(\bar{t}x - \bar{h}) + (1 - 2\alpha)\mathbb{E}_\xi[\zeta^+(x, \xi)] \\ \text{s.t.} & x \in \mathcal{B} \end{array} \right\} \quad (2.128)$$

with  $\alpha \in [0, \frac{1}{2}]$  prescribed. Both problems are clearly convex programming problems.

Interpreting  $\zeta^+(x, \xi)$  as gain (and, consequently,  $\zeta^-(x, \xi)$  as loss), we choose the parameter  $\alpha$  such that  $\alpha \in [\frac{1}{2}, 1]$  holds. By utilizing  $K_\alpha$  with  $\alpha \in [\frac{1}{2}, 1]$ , the corresponding optimization problems are analogous to the two models above, with reversed inequality constraint in (2.127) and with changing “min” to “max” in (2.128).

If the probability distribution of  $\xi$  is finite discrete, problems (2.127) and (2.128) can be equivalently formulated as linear programming problems. This can be done analogously as above for (2.116) and (2.117), we leave the details as an exercise for the reader.

For the case of a finite discrete distribution the feasible set is polyhedral; an analogous representation holds as in Theorem 2.13.. We formulate it for the case  $\alpha \in [0, \frac{1}{2}]$ , the variant with  $\alpha \in [\frac{1}{2}, 1]$  can be obtained from this in a straightforward way. Let  $\widehat{\mathcal{S}}(\alpha) = \{x \mid K_\alpha(x) \leq 0\}$  be the set of feasible solutions corresponding to this type of integrated chance constraint.

**Theorem 2.14.** *Klein Haneveld and Van der Vlerk [191]. For  $\gamma \geq 0$ ,  $\mathcal{S}(\gamma)$  is a polyhedral set and the following representation holds:*

$$\{x \mid K_\alpha(x) \leq 0\} = \bigcap_{\mathcal{H} \subset \mathcal{N}} \{x \mid (1 - 2\alpha) \sum_{k \in \mathcal{H}} p_k(t^k x - h^k) + \alpha(\bar{t}x - \bar{h}) \leq 0\}.$$

*Proof:* The proof runs along the same lines as the proof for Theorem 2.13.. □

### Joint integrated probability functions

Let  $s > 1$  and  $\zeta(x, \xi) = T(\xi)x - h(\xi)$ . Analogously as before, define

$$\zeta^+(x, \xi) := [T(\xi)x - h(\xi)]^+$$

and

$$\zeta^-(x, \xi) := [T(\xi)x - h(\xi)]^-,$$

where on the right-hand-side the positive- and negative parts of the vectors are defined in a componentwise fashion. For a heuristic introduction of joint integrated chance constraints we proceed analogously as in the case  $s = 1$ . We assume that losses are represented by negative values of  $\zeta(x, \xi)$ . The probability constraint in

expected value terms looks as follows (see (2.33) on page 90):

$$\mathbb{P}_\xi(\zeta(x, \xi) \geq 0) \geq \alpha \iff \mathbb{E}_\xi[\chi(\max_{1 \leq i \leq s} \zeta_i^-(x, \xi))] \leq 1 - \alpha.$$

Analogously to the special case  $s = 1$ , dropping  $\chi$  results in the *joint integrated chance constraint* (cf. (2.115)):

$$\mathbb{E}_\xi[\max_{1 \leq i \leq s} \zeta_i^-(x, \xi)] \leq \gamma$$

with prescribed maximal tolerable loss  $\gamma$ . We proceed by defining the quality measures for random variables by

$$\begin{aligned} \rho_{\text{jic}}^+(\vartheta) &:= \mathbb{E}[\max_{1 \leq i \leq s} \vartheta_i^+] \quad \text{and} \\ \rho_{\text{jic}}^-(\vartheta) &:= \mathbb{E}[\max_{1 \leq i \leq s} \vartheta_i^-], \quad \vartheta \in \mathcal{L}_s^1. \end{aligned}$$

This results in the evaluation functions

$$\begin{aligned} K_J(x) &:= \rho_{\text{jic}}^+(\zeta(x, \xi)) = \mathbb{E}_\xi[\max_{1 \leq i \leq s} \zeta_i^+(x, \xi)] \quad \text{and} \\ H_J(x) &:= \rho_{\text{jic}}^-(\zeta(x, \xi)) = \mathbb{E}_\xi[\max_{1 \leq i \leq s} \zeta_i^-(x, \xi)], \end{aligned}$$

respectively. The functions  $K_J(x)$  and  $H_J(x)$  will be called *joint integrated probability functions*.

The attractive property of convexity remains preserved by the generalization:

**Proposition 2.40.** *Both  $H_J(x)$  and  $K_J(x)$  are convex functions on  $\mathbb{R}^n$ .*

*Proof:* The proof is similar to the proof for Proposition 2.39.. For any fixed  $\xi$ ,  $\zeta_i(\cdot, \xi)^+$  is a convex function for the same reasons as in the case  $s = 1$ , see the proof of Proposition 2.39..  $\max_{1 \leq i \leq s} \zeta_i^+(\cdot, \xi)$  is the point-wise maximum of convex functions, consequently this function is also convex for each fixed  $\xi$  (see, for instance, Rockafellar [281]). Taking the expected value w.r. to  $\xi$  preserves convexity, see the proof of Proposition 2.39., therefore  $K_J$  is a convex function. The proof for  $H_J$  is analogous.  $\square$

Let us emphasize that the convexity property holds also for a random technology matrix and without any restriction on the probability distribution of  $\xi$ , beyond the existence of the expected value.

The prototype SLP problems are formulated as follows:

$$\left. \begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & \mathbb{E}_\xi[\max_{1 \leq i \leq s} \zeta_i^+(x, \xi)] \leq \gamma \\ & x \in \mathcal{B} \end{aligned} \right\} \quad (2.129)$$

and

$$\left. \begin{array}{ll} \min & c^T x + \mathbb{E}_{\xi} \left[ \max_{1 \leq i \leq s} \zeta_i^+(x, \xi) \right] \\ \text{s.t.} & x \in \mathcal{B} \end{array} \right\} \quad (2.130)$$

with the prescribed maximal loss level  $\gamma \geq 0$ . Both of them are obviously convex programming problems and due to the convexity of  $K_J$ , the feasible set is closed also for (2.129).

Reversing the inequality in the integrated chance constraint in (2.129) and changing min to max in (2.130) leads in general in both cases to non-convex optimization problems.

There is no change in the behavior of the optimization problems if we utilize  $H_J$  instead of  $K_J$  in the problem formulations.

Assume next that  $\xi$  has a finite discrete distribution, specified in (2.118). Then (2.129) and (2.130) take the form

$$\left. \begin{array}{ll} \min & c^T x \\ \text{s.t.} & \sum_{k=1}^N p_k \max_{1 \leq i \leq s} (t_i^k x - h^k)^+ \leq \gamma \\ & x \in \mathcal{B} \end{array} \right\} \quad (2.131)$$

and

$$\left. \begin{array}{ll} \min & c^T x + \sum_{k=1}^N p_k \max_{1 \leq i \leq s} (t_i^k x - h^k)^+ \\ \text{s.t.} & x \in \mathcal{B}. \end{array} \right\} \quad (2.132)$$

We introduce auxiliary variables  $y_i^k$  and  $z^k$ ,  $i = 1, \dots, s$ ,  $k = 1, \dots, N$  and formulate equivalent linear programming problems to (2.131) and (2.132) as follows (see Klein Haneveld and Van der Vlerk [191]):

$$\left. \begin{array}{ll} \min & c^T x \\ \text{s.t.} & \sum_{k=1}^N p_k z^k \leq \gamma \\ & t_i^k x - y_i^k \leq h_i^k, \quad k = 1, \dots, N, i = 1, \dots, s \\ & -y_i^k + z^k \geq 0, \quad k = 1, \dots, N, i = 1, \dots, s \\ & y_i^k \geq 0, \quad k = 1, \dots, N, i = 1, \dots, s \\ & z^k \geq 0, \quad k = 1, \dots, N \\ & x \in \mathcal{B} \end{array} \right\} \quad (2.133)$$

and

$$\left. \begin{aligned}
& \min c^T x + \sum_{k=1}^N p_k z^k \\
& \text{s.t. } \left. \begin{aligned}
t_i^k x - y_i^k &\leq h_i^k, k=1, \dots, N, i=1, \dots, s \\
-y_i^k + z^k &\geq 0, k=1, \dots, N, i=1, \dots, s \\
y_i^k &\geq 0, k=1, \dots, N, i=1, \dots, s \\
z^k &\geq 0, k=1, \dots, N
\end{aligned} \right\} \quad (2.134) \\
& x \in \mathcal{B}.
\end{aligned} \right\}$$

The equivalence can readily be proved, based on the following fact: if  $\bar{x}$ ,  $\bar{y}_i^k$ ,  $\bar{z}^k$ ,  $k=1, \dots, N$ ,  $i=1, \dots, s$  is a feasible solution of either (2.133) or (2.134) then the following inequality holds:

$$\sum_{k=1}^N p_k \max_{1 \leq i \leq s} (t_i^k \bar{x} - h_i^k)^+ \leq \sum_{k=1}^N p_k \bar{z}^k.$$

Let  $\mathcal{S}_J(\gamma) = \{x \mid \mathbb{E}_\xi [\max_{1 \leq i \leq s} \zeta_i^+(x, \xi)] \leq \gamma\}$ . For  $\xi$  having a finite discrete distribution, Theorem 2.13. has the following generalization:

**Theorem 2.15.** *Klein Haneveld and Van der Vlerk [191]. For  $\gamma \geq 0$ ,  $\mathcal{S}_J(\gamma)$  is a polyhedral set and the following representation holds:*

$$\{x \mid K_J(x) \leq \gamma\} = \bigcap_{\mathcal{H} \subset \mathcal{N}} \bigcap_{l \in \mathcal{J}^{\mathcal{H}}} \{x \mid \sum_{k \in \mathcal{H}} p_k (t_{l_k}^k x - h_{l_k}^k) \leq \gamma\}, \quad (2.135)$$

where  $\mathcal{J} = \{1, \dots, s\}$ ,  $\mathcal{J}^{\mathcal{H}} := \{l := (l_k, k \in \mathcal{H}) \mid l_k \in \mathcal{J} \text{ for all } k \in \mathcal{H}\}$ , and  $t_{l_k}^k$  is the  $l_k^{\text{th}}$  row of  $T^k$ .

*Proof:* We have

$$\begin{aligned}
K_J(x) &= \mathbb{E}_\xi [\max_{1 \leq i \leq s} \zeta_i^+(x, \xi)] = \sum_{k=1}^N p_k \max_{1 \leq i \leq s} \zeta_i^+(x, \hat{\xi}^k) \\
&= \sum_{k \in \mathcal{N}^+} p_k \max_{1 \leq i \leq s} \zeta_i(x, \hat{\xi}^k)
\end{aligned}$$

where  $\mathcal{N}^+ := \{k \in \mathcal{N} \mid \max_{1 \leq i \leq s} \zeta_i(x, \hat{\xi}^k) > 0\}$ . Thus we get for the constraint:

$$\begin{aligned}
K_J(x) \leq \gamma &\iff \sum_{k \in \mathcal{N}^+} p_k \max_{1 \leq i \leq s} \zeta_i(x, \hat{\xi}^k) \leq \gamma \\
&\iff \max_{\mathcal{H} \subset \mathcal{N}} \sum_{k \in \mathcal{H}} p_k \max_{1 \leq i \leq s} \zeta_i(x, \hat{\xi}^k) \leq \gamma \\
&\iff \max_{\mathcal{H} \subset \mathcal{N}} \max_{l \in \mathcal{J}^{\mathcal{H}}} \sum_{k \in \mathcal{H}} p_k \zeta_{l_k}(x, \hat{\xi}^k) \leq \gamma.
\end{aligned}$$



Substituting the definition of  $\zeta_i(x, \hat{\xi}^k)$  and noting that the number of linear inequalities which determine  $\mathcal{S}_J(\gamma)$  is obviously finite yields the result.  $\square$

For counting the inequalities in (2.135) let us observe first that the number of inequalities for a fixed index set  $\mathcal{K}$  (the cardinality of  $\mathcal{S}^{|\mathcal{K}|}$ ) is  $s^{|\mathcal{K}|}$ . Adding up for all subsets of  $\mathcal{N}$  (except of  $\emptyset$ ) results in

$$\sum_{k=1}^N \binom{N}{k} s^k = (s+1)^N - 1.$$

The models in this section are due to Klein Haneveld [188] and have been subsequently investigated by Klein Haneveld and Van der Vlerk [191]. For further properties of integrated chance constraints see these references.

### 2.4.2 A model involving conditional expectation

We consider negative values of the random variable  $\zeta(x, \xi)$  as losses and will discuss constraints which are based on the conditional expectation of the loss given that a loss occurs. This corresponds to the quality measure for random variables

$$\rho_{\text{cexp}}(\vartheta) := \mathbb{E}[-\vartheta \mid \vartheta < 0], \quad \vartheta \in \mathcal{L}_1^1, \quad (2.136)$$

if  $\mathbb{P}(\vartheta < 0) > 0$  and  $\rho_{\text{cexp}}(\vartheta) := 0$  otherwise. Assuming that  $\vartheta$  has a continuous distribution, we have the following close relation between  $\rho_{\text{cexp}}$  and the quality measure  $\rho_{\text{sic}}^-$  which lead to integrated chance constraints (see (2.114) on page 143):  $\rho_{\text{sic}}^-(\vartheta) = \rho_{\text{cexp}}(\vartheta)\mathbb{P}(\vartheta < 0)$ . This follows immediately from

$$\mathbb{E}[\vartheta^-] = - \int_{-\infty}^0 t dF_{\vartheta}(t) = \mathbb{E}[-\vartheta \mid \vartheta < 0] \cdot \mathbb{P}(\vartheta < 0).$$

Constraints of the form

$$\mathbb{E}_{\xi}[-\zeta(x, \xi) \mid \zeta(x, \xi) < 0] \leq \gamma \quad (2.137)$$

will be considered, with  $\gamma$  being a prescribed upper bound for the conditional loss size. In general, constraints of this type result in non-convex optimization problems. In the special case when only the right-hand-side is stochastic, the feasible set corresponding to the constraint (2.137) is convex for a broad class of univariate distributions. We choose

$$\zeta(x, \xi) = t^T x - \xi,$$

where  $t \in \mathbb{R}^n$  is a deterministic vector and  $\xi$  is a random variable.

The following result will be utilized:

**Proposition 2.41.** *Assume that  $\xi$  has a continuous distribution with a logconcave density function. Assume furthermore that the expected value of  $\xi$  exists. Then*

$$l(t) := \mathbb{E}[\xi - t \mid \xi - t > 0]$$

*is a monotonically decreasing function of  $t$ .*

*Proof:* This is a well-known fact in reliability theory where  $l(t)$  is called mean residual life. For a proof see, for instance, Prékopa [266].  $\square$

We assume that for  $\xi$  the assumptions of the theorem hold. Then (2.137) takes the form

$$l(t^T x) \leq \gamma \iff t^T x \geq l^{-1}(\gamma),$$

where  $l^{-1}$  is to be understood as a generalized inverse defined as follows  $l^{-1}(z) := \inf\{z \mid l(z) \leq \gamma\}$ . Consequently, the constraint (2.137) can be reformulated as a deterministic linear constraint.

### 2.4.3 Conditional Value at Risk

We assume that positive values of  $\zeta(x, \xi)$  represent losses. For motivating the quality (risk) measure which will be introduced, let us start with computing a conditional expected value. Let  $\vartheta$  be a random variable with finite expected value,  $F_\vartheta$  its distribution function,  $0 < \alpha < 1$ , and  $v_\alpha$  an  $\alpha$ -quantile of the distribution of  $\vartheta$  (see Section 2.3 for the definition of quantiles). Note that due to  $\alpha < 1$ ,  $\mathbb{P}(\vartheta < v_\alpha) < 1$  holds and consequently we have  $\mathbb{P}(\vartheta \geq v_\alpha) > 0$ . Introducing the notation  $\pi_\alpha = \mathbb{P}(\vartheta \geq v_\alpha)$  we get

$$\begin{aligned} \mathbb{E}[\vartheta \mid \vartheta \geq v_\alpha] &= \frac{1}{\mathbb{P}(\vartheta \geq v_\alpha)} \int_{v_\alpha}^{\infty} t dF_\vartheta(t) \\ &= \frac{1}{\pi_\alpha} \left[ \int_{v_\alpha}^{\infty} t dF_\vartheta(t) - v_\alpha \int_{v_\alpha}^{\infty} dF_\vartheta(t) + v_\alpha \pi_\alpha \right] \\ &= \frac{1}{\pi_\alpha} \left[ \int_{v_\alpha}^{\infty} (t - v_\alpha) dF_\vartheta(t) + v_\alpha \pi_\alpha \right] \\ &= v_\alpha + \frac{1}{\pi_\alpha} \int_{-\infty}^{\infty} (t - v_\alpha)^+ dF_\vartheta(t) \\ &= v_\alpha + \frac{1}{\mathbb{P}(\vartheta \geq v_\alpha)} \mathbb{E}[(\vartheta - v_\alpha)^+]. \end{aligned} \tag{2.138}$$

If  $F_{\vartheta}$  is continuous, we have  $F_{\vartheta}(v_{\alpha}) = \alpha$  and  $\mathbb{P}(\vartheta \geq v_{\alpha}) = 1 - \alpha$ . Consequently, in this case the above relation takes the form

$$\mathbb{E}[\vartheta \mid \vartheta \geq v_{\alpha}] = v_{\alpha} + \frac{1}{1 - \alpha} \mathbb{E}[(\vartheta - v_{\alpha})^+]. \quad (2.139)$$

On the other hand, due to a well-known fact in probability theory, the following optimization problem has a solution for any  $0 < \alpha < 1$  and the solution set is the interval of  $\alpha$ -quantiles:

$$\min_z (\alpha \mathbb{E}[(\vartheta - z)^+] + (1 - \alpha) \mathbb{E}[(\vartheta - z)^-]). \quad (2.140)$$

Using  $z = z^+ - z^-$ , we have

$$\alpha(\vartheta - z)^+ + (1 - \alpha)(\vartheta - z)^- = (1 - \alpha) \left[ z - \vartheta + \frac{1}{1 - \alpha} (\vartheta - z)^+ \right].$$

Taking expectation, this leads to the equivalent formulation of the objective function of the unconstrained minimization problem (2.140) as

$$(1 - \alpha) \left( z + \frac{1}{(1 - \alpha)} \mathbb{E}[(\vartheta - z)^+] \right) - (1 - \alpha) \mathbb{E}[\vartheta]$$

which results in the following equivalent formulation of (2.140):

$$\min_z \left( z + \frac{1}{(1 - \alpha)} \mathbb{E}[(\vartheta - z)^+] \right), \quad (2.141)$$

see Rockafellar and Uryasev [282]. Utilizing (2.38.) and introducing the notation  $u_c(z)$  for the objective function of this unconstrained optimization problem, we have

$$u_c(z) := z + \frac{1}{(1 - \alpha)} \mathbb{E}[(\vartheta - z)^+] = z + \frac{1}{(1 - \alpha)} \int_z^{\infty} (1 - F_{\vartheta}(t)) dt. \quad (2.142)$$

The function  $u_c(\cdot)$  is obviously convex. In fact, for each fixed  $\vartheta$ ,  $(\vartheta - z)^+$  is obtained from the convex function  $(\cdot)^+$  by substituting a linear function, therefore it is convex. Taking the expected value clearly preserves convexity, see, for instance, the proof of Proposition 2.39. on page 143. Thus (2.141) is a convex programming problem. As mentioned above, the set of solutions of (2.141) consists of the set of  $\alpha$ -quantiles of the distribution of  $\vartheta$ . This is easy to see under the assumption that  $F_{\vartheta}$  is continuous. In fact, due to the integral representation in (2.142) it follows immediately that  $u_c(z)$  is continuously differentiable. Taking into account that (2.141) is a convex programming problem, the set of optimal solutions is determined by the equation  $\frac{du_c(z)}{dz} = 0$  which can be written as

$$1 - \frac{1}{(1 - \alpha)} (1 - F_{\vartheta}(z)) = 0 \iff F_{\vartheta}(z) = \alpha,$$

which obviously has as solution set the interval of  $\alpha$ -quantiles. Based on the fact that for  $u_c(z)$ , being a convex function, the left- and right-sided derivatives exist for all  $z \in \mathbb{R}$ , a proof for the general case can be found in Rockafellar and Uryasev [283]. An elementary proof is given by Pflug [254].

The solution set of problem (2.141) being the interval of  $\alpha$ -quantiles, it follows that in particular the Value at Risk  $v_\alpha := v(\vartheta, \alpha)$  (for the definition see (2.101) on page 137) is an optimal solution of (2.140). Taking into account (2.139) it follows that for continuous  $F_\vartheta$  the optimal objective value in (2.141) is  $\mathbb{E}[\vartheta \mid \vartheta \geq v(\vartheta, \alpha)]$ . Consequently, in this case, the optimal objective value of (2.141) is the conditional expected value of the loss, given that the loss is greater than or equal to VaR. This motivates introducing the following risk measure for random variables:

$$\rho_{\text{CVaR}}^\alpha(\vartheta) := v_c(\vartheta, \alpha) := \min_z \left[ z + \frac{1}{1-\alpha} \mathbb{E}[(\vartheta - z)^+] \right], \quad \vartheta \in \mathcal{L}_1^1, \quad (2.143)$$

defined on the linear space of random variables with finite expected value. For the case when  $F_\vartheta$  is continuous, we have

$$v_c(\vartheta, \alpha) = \mathbb{E}[\vartheta \mid \vartheta \geq v(\vartheta, \alpha)], \quad (2.144)$$

where  $v(\vartheta, \alpha)$  is the Value at Risk corresponding to  $\vartheta$  and  $\alpha$ , see (2.101) on page 137.

Because the Value at Risk  $v(\vartheta, \alpha)$  is an optimal solution of (2.140), substituting it for  $z$  in (2.140) immediately leads to the inequality

$$v_c(\vartheta, \alpha) \geq v(\vartheta, \alpha).$$

The risk measure  $v_c(\vartheta, \alpha)$  has been introduced by Rockafellar and Uryasev [282] for a financial application where the authors call it Conditional Value-at-Risk (CVaR). We will use this terminology also in our context. For the corresponding evaluation function for  $x$  we consider the random variable

$$\zeta(x, \xi) := t(\xi)^T x - h(\xi),$$

where  $t(\xi)$  is an  $n$ -dimensional random vector and  $h(\xi)$  is a random variable. The evaluation function, denoted by  $v_c(x, \alpha)$ , is

$$v_c(x, \alpha) := \min_z \left[ z + \frac{1}{1-\alpha} \mathbb{E}[(\zeta(x, \xi) - z)^+] \right].$$

Introducing the notation

$$w_c^\alpha(x, z) := z + \frac{1}{1-\alpha} \mathbb{E}[(\zeta(x, \xi) - z)^+]$$

we have the shorthand form

$$v_c(x, \alpha) = \min_z w_c^\alpha(x, z).$$

Let  $\Psi(x, \cdot)$  denote the probability distribution function of  $\zeta(x, \xi)$ . For later reference we formulate the specialization of the general findings above for the case  $\vartheta = \zeta(x, \xi)$  as a separate proposition:

**Proposition 2.42.** *Let  $x \in \mathbb{R}^n$  be arbitrary and assume  $0 < \alpha < 1$ . For the unconstrained optimization problem*

$$\min_z \left[ z + \frac{1}{1-\alpha} \mathbb{E}[(\zeta(x, \xi) - z)^+] \right]$$

*the following assertions hold: this is a convex optimization problem; the optimal solution exists and is attained; the set of optimal solutions coincides with the set of  $\alpha$ -quantiles of  $\Psi(x, \cdot)$ .*

*Proof:* The proof follows readily from the general case.  $\square$

Let us consider  $w_c^\alpha(x, z)$  as a function in the joint variables  $(x, z)$ .

**Proposition 2.43.**  *$w_c^\alpha$  is a convex function.*

*Proof:*  $\zeta(x, \xi) - z$  being a linear-affine function of  $(x, z)$  and  $(\cdot)^+$  being a convex function implies that the composite function  $(\zeta(x, \xi) - z)^+$  is jointly convex in  $(x, z)$  for each fixed  $\xi$ . Proceeding analogously as in the proof of Proposition 2.39. on page 143 it is easy to see that taking the expected value preserves convexity.  $\square$

Next we formulate the corresponding optimization problems. SLP models involving CVaR in the objective can be formulated as follows:

$$\begin{aligned} \min_x c^T x + \min_z \left[ z + \frac{1}{1-\alpha} \mathbb{E}[(\zeta(x, \xi) - z)^+] \right] \\ \text{s.t.} \quad x \in \mathcal{B}. \end{aligned} \quad (2.145)$$

This can obviously be written in the equivalent form

$$\begin{aligned} \min_{(x, z)} c^T x + z + \frac{1}{1-\alpha} \mathbb{E}[(\zeta(x, \xi) - z)^+] \\ \text{s.t.} \quad x \in \mathcal{B}. \end{aligned} \quad (2.146)$$

The equivalence is due to the fact, that for each fixed  $x \in \mathcal{B}$  in (2.146) it is sufficient to take into account the corresponding  $z$  for which the sum of the second and third terms in the objective is minimal with fixed  $x$  (this minimum is attained for any  $x$ , see the discussion above).

Proposition 2.43. immediately implies that (2.146) is a convex programming problem for arbitrary probability distribution of  $\xi$ . Let  $(x^*, z^*)$  be an optimal solution of (2.146). Then  $z^*$  is an  $\alpha$ -quantile of  $\Psi(x^*, \cdot)$  and we have

$$\begin{aligned} v(x^*, \alpha) &\leq z^* \\ v_c(x^*, \alpha) &= z^* + \frac{1}{1-\alpha} \mathbb{E}[(\zeta(x^*, \xi) - z^*)^+], \end{aligned}$$

where  $v(x^*, \alpha)$  is the Value at Risk corresponding to  $x^*$  (for VaR see (2.103) on page 137).

Let us turn our attention to the particular case when  $\xi$  has a finite discrete distribution with  $N$  realizations and corresponding probabilities given as

$$\begin{pmatrix} p_1 & \dots & p_N \\ \widehat{\xi}^1 & \dots & \widehat{\xi}^N \end{pmatrix} \quad (2.147)$$

with  $p_i > 0 \forall i$  and  $\sum_{i=1}^N p_i = 1$ . Let us introduce the notation  $t^k := t(\widehat{\xi}^k)$ ,  $h^k := h(\widehat{\xi}^k)$ ,  $k = 1, \dots, N$ . The optimization problem (2.146) specializes as follows:

$$\left. \begin{array}{ll} \min_{(x,z)} & c^T x + z + \frac{1}{1-\alpha} \sum_{k=1}^N p_k (\zeta(x, \widehat{\xi}^k) - z)^+ \\ \text{s.t.} & x \in \mathcal{B}. \end{array} \right\} \quad (2.148)$$

Using a well-known idea in optimization, this nonlinear programming problem can be transformed into a linear programming problem by introducing additional variables  $y_k$  for representing  $(\zeta(x, \widehat{\xi}^k) - z)^+$  for all  $k = 1, \dots, N$ . The equivalent linear programming problem is the following:

$$\left. \begin{array}{ll} \min_{(x,z,y)} & c^T x + z + \frac{1}{1-\alpha} \sum_{k=1}^N p_k y_k \\ \text{s.t.} & \zeta(x, \widehat{\xi}^k) - z - y_k \leq 0, \quad k = 1, \dots, N \\ & y_k \geq 0, \quad k = 1, \dots, N \\ & x \in \mathcal{B}. \end{array} \right\} \quad (2.149)$$

The equivalence can be seen as follows. If  $(\bar{x}, \bar{z})$  is a feasible solution of (2.148) then taking  $\bar{y}_k = (\zeta(\bar{x}, \widehat{\xi}^k) - \bar{z})^+$  for all  $k$ , the resulting  $(\bar{x}, \bar{z}, \bar{y}_k, k = 1, \dots, N)$  is obviously feasible in (2.149) with equal objective values. Vice versa, let  $(\hat{x}, \hat{z}, \hat{y}_k, k = 1, \dots, N)$  be a feasible solution of (2.149). Then  $(\hat{x}, \hat{z})$  is evidently feasible in (2.148) and due to the first constraint in (2.149), the corresponding objective value in (2.148) does not exceed the objective value in (2.149). This proves the equivalence. Substituting for  $\zeta(x, \widehat{\xi}^k)$  results in the final form of the equivalent LP problem:

$$\left. \begin{array}{ll} \min_{(x,z,y)} & c^T x + z + \frac{1}{1-\alpha} \sum_{k=1}^N p_k y_k \\ \text{s.t.} & t^k x - z - y_k \leq h^k, \quad k = 1, \dots, N \\ & y_k \geq 0, \quad k = 1, \dots, N \\ & x \in \mathcal{B}. \end{array} \right\} \quad (2.150)$$

Let us turn our attention to the optimization problems with CVaR constraints

$$\left. \begin{array}{l} \min_x c^T x \\ \text{s.t. } v_c(x, \alpha) \leq \gamma \\ x \in \mathcal{B}, \end{array} \right\} \quad (2.151)$$

where  $\gamma$  is a prescribed threshold. Substituting for  $v_c(x, \alpha)$  results in

$$\left. \begin{array}{l} \min_x c^T x \\ \text{s.t. } \min_z w_c^\alpha(x, z) \leq \gamma \\ x \in \mathcal{B}. \end{array} \right\} \quad (2.152)$$

Due to Proposition 2.42, the minimum in the first inequality is attained for any  $x \in \mathcal{B}$ . Therefore, for any fixed  $x$ , the first inequality holds if and only if there exists a  $z \in \mathbb{R}$  for which  $w_c^\alpha(x, z) \leq \gamma$  holds. Substituting for  $w_c^\alpha(x, z)$ , the following equivalent formulation results:

$$\left. \begin{array}{l} \min_{(x,z)} c^T x \\ \text{s.t. } z + \frac{1}{1-\alpha} \mathbb{E}[(\zeta(x, \xi) - z)^+] \leq \gamma \\ x \in \mathcal{B}. \end{array} \right\} \quad (2.153)$$

This is a nonlinear optimization problem involving a nonlinear constraint. From Proposition 2.43, immediately follows that this problem belongs to the class of convex optimization problems, for an arbitrary probability distribution of  $\xi$ .

Let us consider the case of a finite discrete distribution of  $\xi$ , as specified in (2.147). This leads to the specialized form

$$\left. \begin{array}{l} \min_{(x,z)} c^T x \\ \text{s.t. } z + \frac{1}{1-\alpha} \sum_{k=1}^N p_k (\zeta(x, \hat{\xi}^k) - z)^+ \leq \gamma \\ x \in \mathcal{B}. \end{array} \right\} \quad (2.154)$$

Using the same transformation as for deriving (2.149), we get the equivalent formulation as

$$\left. \begin{array}{l} \min_{(x,z,y)} c^T x \\ \text{s.t. } z + \frac{1}{1-\alpha} \sum_{k=1}^N p_k y_k \leq \gamma \\ \zeta(x, \hat{\xi}^k) - z - y_k \leq 0, \quad k = 1, \dots, N \\ y_k \geq 0, \quad k = 1, \dots, N \\ x \in \mathcal{B}. \end{array} \right\} \quad (2.155)$$

The final equivalent form is obtained by substituting for  $\zeta(x, \hat{\xi}^k)$ :

$$\left. \begin{array}{ll} \min_{(x,z,y)} & c^T x \\ \text{s.t.} & z + \frac{1}{1-\alpha} \sum_{k=1}^N p_k y_k \leq \gamma \\ & t^k x - z - y_k \leq h^k, \quad k = 1, \dots, N \\ & y_k \geq 0, \quad k = 1, \dots, N \\ & x \in \mathcal{B}. \end{array} \right\} \quad (2.156)$$

Finally let us discuss the interpretation of  $v_c(x, \alpha)$ ,  $0 < \alpha < 1$ . From our general discussions at the beginning of this section it follows readily, that under the assumption that the probability distribution function  $\Psi(x, \cdot)$  of the random variable  $\zeta(x, \xi)$  is continuous, we have the relation

$$v_c(x, \alpha) = \mathbb{E}[\zeta(x, \xi) \mid \zeta(x, \xi) \geq v(x, \alpha)],$$

where  $v(x, \alpha)$  is the Value at Risk corresponding to  $\zeta(x, \xi)$  and  $\alpha$ . If  $\xi$  has for example a finite discrete distribution then this relation does not hold anymore in general.

For the following discussion let us consider again a random variable  $\vartheta$  and assume that the expected value exists. In this terms the above relation has been formulated in (2.144) under the assumption that the distribution function  $F_\vartheta$  of  $\vartheta$  is continuous. It has the form

$$\rho_{\text{CVaR}}^\alpha(\vartheta) = \mathbb{E}[\vartheta \mid \vartheta \geq v(\vartheta, \alpha)],$$

where  $v(\vartheta, \alpha)$  is the VaR corresponding to  $\vartheta$  and  $\alpha$ , see (2.101) on page 137.

For general distributions an interpretation has been given by Rockafellar and Uryasev in [283]. The conditional expectation relation above holds in general, if the original distribution function  $F_\vartheta$  is replaced by the upper-tail distribution function  $F_\vartheta^\alpha$  defined as follows:

$$F_\vartheta^\alpha(y) = \begin{cases} 0 & \text{if } y < v(\vartheta, \alpha) \\ \frac{F_\vartheta(y) - \alpha}{1 - \alpha} & \text{if } y \geq v(\vartheta, \alpha). \end{cases}$$

Another interpretation, representing  $\alpha$ -CVaR as a mean over  $\alpha$  of  $\alpha$ -VaR, has been found by Acerbi [3]. For further properties of CVaR see Rockafellar and Uryasev [283] and Acerbi and Tasche [4]. In the latter paper several related risk measures and their interrelations are also presented.

## Exercises

**2.8.** With integrated chance constraints of the second kind the SLP-problems have been formulated on page 147 as (2.127) and (2.128), with  $\alpha \in [0, \frac{1}{2}]$  prescribed.



Formulate the equivalent linear programming problems for the case when  $\xi$  has a finite discrete distribution.

**2.9.** Consider the following pair of SLP problems:

$$\left. \begin{array}{ll} \min_x & x_1 + 2x_2 \\ \text{s.t.} & \mathbb{E}[\zeta^+(x, \xi)] \leq 0.8 \\ & x_1 + x_2 \geq 3 \\ & x_1, \quad x_2 \geq 0 \end{array} \right\} (ICC)$$

and

$$\left. \begin{array}{ll} \min_{x,z} & x_1 + 2x_2 \\ \text{s.t.} & \rho_{CVaR}^\alpha(\zeta(x, \xi)) \leq 0.8 \\ & x_1 + x_2 \geq 3 \\ & x_1, \quad x_2 \geq 0 \end{array} \right\} (CVaR)$$

with  $\zeta(x, \xi) := \xi_1 x_1 + \xi_2 x_2 - 0.5$ ,  $\alpha = 0.95$  and  $(\xi_1, \xi_2)$  having a finite discrete distribution with three realizations, given by the realizations tableau:

$p_k$	0.2	0.5	0.3
$\xi_1^k$	1	-0.5	0.5
$\xi_2^k$	2	0.5	-0.5

Notice that both problems are set up on the same data-set; (ICC) is formulated with an integrated chance constraint whereas (CVaR) is an SLP model with a CVaR-constraint.

- Solve both problems by utilizing SLP-IOR and compare the optimal objective values  $z_{ICC}^*$  and  $z_{CVaR}^*$ .
- You will see that  $z_{ICC}^* < z_{CVaR}^*$  holds. This is a typical result when setting up and solving both type of problems based on the same data-set. Give an explanation for this phenomenon.

## 2.5 Models built with deviation measures

In this section we deal exclusively with quality measures expressing risk. Similarly as in Section 2.2.3, for the sake of simplicity we employ the notation  $\eta := t(\xi)$  and replace the right-hand-side  $h(\xi)$  by  $\xi$ . Thus we consider the random variable

$$\zeta(x, \eta, \xi) := \eta^T x - \xi,$$

where  $\eta$  denotes an  $n$ -dimensional random vector and  $\xi$  is a random variable. We will assume in this section that the expected value of  $(\eta^T, \xi)$  exists and will use the notation  $\mu := \mathbb{E}[\eta] \in \mathbb{R}^n$  and  $\mu_{n+1} := \mathbb{E}[\xi] \in \mathbb{R}$ .

### 2.5.1 Quadratic deviation

The risk measure is chosen as

$$\rho_Q(\vartheta) := \sqrt{\mathbb{E}[\vartheta^2]} = \sqrt{\text{Var}[\vartheta] + (\mathbb{E}[\vartheta])^2}, \quad \vartheta \in \mathcal{L}_1^2,$$

defined on the linear space of random variables with finite variance. We assume that the second moments for the random vector  $(\eta^T, \xi)$  exist and the distribution is non-degenerate, meaning that the covariance matrix of this random vector is positive definite. The corresponding evaluation function will be denoted by  $Q(x)$  and has the form

$$Q(x) := \sqrt{\mathbb{E}[(\eta^T x - \xi)^2]} = \sqrt{\text{Var}[\eta^T x - \xi] + (\mu^T x - \mu_{n+1})^2}. \quad (2.157)$$

It is interpreted as measuring the deviation between the random variables  $\eta^T x$  and  $\xi$ .

**Proposition 2.44.**  *$Q$  is a convex function.*

*Proof:* An elegant proof of this assertion can be obtained by combining Propositions 2.47. and 2.50. in Section 2.7. Here we present a direct elementary proof. Let us consider the functions  $q : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  and  $\hat{q} : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  defined as  $q(x, x_{n+1}) = \mathbb{E}[(\eta^T x + \xi x_{n+1})^2]$  and  $\hat{q}(x, x_{n+1}) = \sqrt{q(x, x_{n+1})}$ , respectively. We will prove that  $\hat{q}$  is a convex function. Due to the relation  $Q(x) = \hat{q}(x, -1)$ , the assertion follows from this.

We consider  $q(x, x_{n+1})$  first. This function is obviously nonnegative,  $q(x, x_{n+1}) \geq 0$  holds for all  $x \in \mathbb{R}^n, x_{n+1} \in \mathbb{R}$ . The function is quadratic

$$\begin{aligned} \mathbb{E}[(\eta^T x + \xi x_{n+1})^2] &= \mathbb{E} \left[ (x^T, x_{n+1}) \begin{pmatrix} \eta \\ \xi \end{pmatrix} (\eta^T, \xi) \begin{pmatrix} x \\ x_{n+1} \end{pmatrix} \right] \\ &= (x^T, x_{n+1}) \begin{pmatrix} \mathbb{E}[\eta \eta^T] & \mathbb{E}[\xi \eta^T] \\ \mathbb{E}[\xi \eta^T] & \mathbb{E}[\xi^2] \end{pmatrix} \begin{pmatrix} x \\ x_{n+1} \end{pmatrix} \end{aligned} \quad (2.158)$$

therefore, because of the nonnegativity of  $q$ , the symmetric matrix in the second line in (2.158) is positive semidefinite. Thus  $q$  is a convex function. In general, the square root of a convex function need not to be convex (take  $z$  and  $\sqrt{z}$  for  $z \geq 0$ ). In our case  $\hat{q}(x, x_{n+1}) = \sqrt{q(x, x_{n+1})}$  is the square root of a positive semidefinite quadratic form, therefore it is convex. To see this let  $D$  be an  $(n \times n)$  symmetric positive semidefinite matrix, we shall prove that  $d(x) := \sqrt{x^T D x}$  is a convex function. For this function  $d(\lambda x) = \lambda d(x)$  holds obviously for all  $\lambda \geq 0$  and  $x \in \mathbb{R}^n$ . Therefore, for proving the

convexity of  $d$ , it is sufficient to prove that  $d(x+y) \leq d(x) + d(y)$  (subadditivity) holds for all  $x, y \in \mathbb{R}^n$  (see Proposition 2.46. on page 174). We have

$$d^2(x+y) = (x+y)^T D(x+y) = x^T D(x+y) + y^T D(x+y). \quad (2.159)$$

By applying the Cauchy–Schwarz inequality to the first term on the right–hand–side we get

$$x^T D(x+y) = [x^T D^{\frac{1}{2}}][D^{\frac{1}{2}}(x+y)] \leq \sqrt{x^T D x} \sqrt{(x+y)^T D(x+y)},$$

where  $D^{1/2}$  denotes the symmetric square root of the positive semidefinite matrix  $D$ . The latter is defined as follows: take the spectral decomposition  $D = T \Lambda T^T$  of  $D$ , where the columns of  $T$  consist of an orthonormal system of eigenvectors of  $D$  and the diagonal elements of the diagonal matrix  $\Lambda$  are the corresponding eigenvalues. Taking  $D^{1/2} := T \Lambda^{1/2} T^T$  we obviously have  $D = D^{1/2} D^{1/2}$ . Performing analogously with the second term in (2.159) and substituting into (2.159) yields the result.  $\square$

Applying (2.158) for  $\hat{\eta}$  and  $\hat{\xi}$ , defined as  $\hat{\eta} := \eta - \mu$  and  $\hat{\xi} := \xi - \mu_{n+1}$ , and setting  $x_{n+1} = -1$  from (2.158) it follows that

$$\mathbb{W}\mathbf{ar}[\eta^T x - \xi] = x^T V x - 2d^T x + v$$

holds, where  $V := \mathbb{E}[\hat{\eta} \hat{\eta}^T] = \mathbb{C}\mathbf{ov}[\eta, \eta]$  is the covariance matrix of  $\eta$ ,  $d := \mathbb{E}[\hat{\eta} \hat{\xi}] = \mathbb{C}\mathbf{ov}[\eta, \xi]$  is the cross-covariance vector between  $\eta$  and  $\xi$ , and  $v := \mathbb{E}[\hat{\xi}^2] = \mathbb{W}\mathbf{ar}[\xi]$  is the variance of  $\xi$ . Note that  $V$  is a positive semidefinite matrix. Thus we have have derived the formula

$$Q(x) = \sqrt{x^T V x - 2d^T x + v + (\mu^T x - \mu_{n+1})^2}.$$

We obtain the following convex optimization problems

$$\left. \begin{array}{l} \min c^T x \\ \text{s.t. } \sqrt{x^T V x - 2d^T x + v + (\mu^T x - \mu_{n+1})^2} \leq \sqrt{\kappa} \\ x \in \mathcal{B} \end{array} \right\} \quad (2.160)$$

and

$$\left. \begin{array}{l} \min \sqrt{x^T V x - 2d^T x + v + (\mu^T x - \mu_{n+1})^2} \\ \text{s.t. } x \in \mathcal{B}. \end{array} \right\} \quad (2.161)$$

Due to the definition of  $Q$ , the expression under the square root is nonnegative for all  $x \in \mathbb{R}^n$ , and the positive square root function is strictly monotonically increasing, consequently we have the equivalent formulation

$$\left. \begin{array}{l} \min c^T x \\ \text{s.t. } x^T V x + (\mu^T x - \mu_{n+1})^2 - 2d^T x \leq \kappa - v \\ x \in \mathcal{B} \end{array} \right\} \quad (2.162)$$

and

$$\begin{aligned} \min & x^T V x + (\mu^T x - \mu_{n+1})^2 - 2d^T x \\ \text{s.t. } & x \in \mathcal{B}. \end{aligned} \quad (2.163)$$

The matrix  $V$  is positive semidefinite, therefore both problems are convex optimization problems. Note that (2.163) is a convex quadratic optimization problem.

A widely used variant of the above risk measure for random variables is the standard deviation:

$$\rho_{\text{Std}}(\vartheta) := \sigma(\vartheta) := \sqrt{\mathbb{E}[(\vartheta - \mathbb{E}[\vartheta])^2]}, \quad \vartheta \in \mathcal{L}_1^2.$$

The evaluation function becomes

$$Q_d(x) = \sqrt{x^T V x - 2d^T x + v} \quad (2.164)$$

leading to the convex optimization problems

$$\begin{aligned} \min & c^T x \\ \text{s.t. } & x^T V x - 2d^T x \leq \kappa - v \\ & x \in \mathcal{B} \end{aligned} \quad (2.165)$$

and

$$\begin{aligned} \min & x^T V x - 2d^T x \\ \text{s.t. } & x \in \mathcal{B}. \end{aligned} \quad (2.166)$$

An important special case is  $\xi \equiv 0$ . The evaluation function becomes

$$\sigma(x) := \sqrt{\mathbb{E}[(\eta^T x - \mu^T x)^2]} = \sqrt{x^T V x}, \quad (2.167)$$

where  $V$  is covariance matrix of  $\eta$ . Because of their practical importance we formulate also the resulting optimization problems. For obvious reasons, these can equivalently be formulated in terms of  $\sigma^2(x)$  as follows:

$$\begin{aligned} \min & c^T x \\ \text{s.t. } & x^T V x \leq \kappa \\ & x \in \mathcal{B} \end{aligned} \quad (2.168)$$

and

$$\begin{aligned} \min & x^T V x \\ \text{s.t. } & x \in \mathcal{B}. \end{aligned} \quad (2.169)$$

Optimization problems of this type are widely used in financial portfolio optimization, see Markowitz [217] and Elton et al. [85].

### 2.5.2 Absolute deviation

Let the risk measure be

$$\rho_A(\vartheta) := \mathbb{E}[|\vartheta|], \quad \vartheta \in \mathcal{L}_1^1, \quad (2.170)$$

defined on the linear space of random variables with finite expected value. Assuming that the expected value of  $(\eta^T, \xi)$  exists we get the corresponding evaluation function

$$A(x) := \mathbb{E}[|\eta^T x - \xi|], \quad (2.171)$$

which is interpreted as measuring deviation between the random variables  $\eta^T x$  and  $\xi$ . Let  $\zeta(x, \eta, \xi) := \eta^T x - \xi$ . We have:

**Proposition 2.45.**  *$A(\cdot)$  is a convex function.*

*Proof:* The absolute-value function being convex and  $\zeta(\cdot, \eta, \xi)$  being linear, the composite function  $|\zeta(\cdot, \eta, \xi)|$  is convex for any fixed realization of  $(\eta, \xi)$ . Taking expected value preserves convexity. The full proof runs analogously as the proof of Proposition 2.39. on page 143.  $\square$

Thus the optimization problems

$$\left. \begin{array}{ll} \min & c^T x \\ \text{s.t.} & \mathbb{E}[|\eta^T x - \xi|] \leq \kappa \\ & x \in \mathcal{B} \end{array} \right\} \quad (2.172)$$

and

$$\left. \begin{array}{ll} \min & \mathbb{E}[|\eta^T x - \xi|] \\ \text{s.t.} & x \in \mathcal{B} \end{array} \right\} \quad (2.173)$$

are convex optimization problems for arbitrary random variables with finite expected value.

The model (2.173) is closely related to simple recourse problems. To see this let  $\eta \equiv t$  with  $t$  being an  $n$ -dimensional deterministic vector and let us formulate this problem equivalently as follows. We introduce the nonnegative random variables  $y$  and  $z$  and make use of the relations  $|u| = u^+ + u^-$  and  $u = u^+ - u^-$  which hold for any real number  $u$ . This results in the following equivalent simple recourse formulation of (2.173)

$$\left. \begin{array}{ll} \min & \mathbb{E}[y+z] \\ \text{s.t.} & t^T x - \xi - y + z = 0 \\ & y \geq 0 \\ & z \geq 0 \\ & x \in \mathcal{B}, \end{array} \right\} \quad (2.174)$$

where the constraints involving random variables are interpreted as usual: they should hold in the almost sure sense. For proving the equivalence let  $\hat{x}$ , along with

the random variables  $\hat{y}$  and  $\hat{z}$  be a feasible solution of (2.174). Then  $\hat{x}$  is obviously a feasible solution of (2.173) and for the corresponding objective function values we have

$$\mathbb{E}[\hat{y} + \hat{z}] \geq \mathbb{E}[|\hat{y} - \hat{z}|] = \mathbb{E}[|t^T \hat{x} - \xi|].$$

In the reverse direction, when  $\bar{x}$  is a feasible solution of (2.173) then setting  $\bar{y} = (t^T \bar{x} - \xi)^+$  and  $\bar{z} = (t^T \bar{x} - \xi)^-$  we get a feasible solution to (2.174) and the objective values are equal. This proves the equivalence.

Let us consider the case of a finite discrete distribution next. Assume that  $(\eta^T, \xi)$  has  $N$  distinct realizations with corresponding probabilities given in the table:

$$\begin{pmatrix} p_1 & \dots & p_N \\ \hat{\eta}^1 & \dots & \hat{\eta}^N \\ \hat{\xi}^1 & \dots & \hat{\xi}^N \end{pmatrix} \quad (2.175)$$

with  $p_i > 0 \forall i$  and  $\sum_{i=1}^N p_i = 1$ . Let  $t^k := (\hat{\eta}^k)^T$ ,  $h^k := \hat{\xi}^k$ ,  $k = 1, \dots, N$ . In this case our optimization problems have the form

$$\left. \begin{array}{ll} \min & c^T x \\ \text{s.t.} & \sum_{k=1}^N p_k |t^k x - h^k| \leq \kappa \\ & x \in \mathcal{B} \end{array} \right\} \quad (2.176)$$

and

$$\left. \begin{array}{ll} \min & \sum_{k=1}^N p_k |t^k x - h^k| \\ \text{s.t.} & x \in \mathcal{B}. \end{array} \right\} \quad (2.177)$$

Both of these problems are nonlinear programming problems. We utilize a transformation, analogous to the transformation which leads to the formulation (2.174). Introducing this time deterministic auxiliary variables  $y_k$  and  $z_k$ ,  $k = 1, \dots, N$ , we obtain the equivalent deterministic linear programming formulations

$$\left. \begin{array}{ll} \min & c^T x \\ \text{s.t.} & \sum_{k=1}^N p_k (y_k + z_k) \leq \kappa \\ & t^k x - y_k + z_k = h^k, \quad k = 1, \dots, N \\ & y_k \geq 0, \quad k = 1, \dots, N \\ & z_k \geq 0, \quad k = 1, \dots, N \\ & x \in \mathcal{B} \end{array} \right\} \quad (2.178)$$

and

$$\left. \begin{aligned} \min \quad & \sum_{k=1}^N p_k(y_k + z_k) \\ \text{s.t.} \quad & t^k x - y_k + z_k = h^k, \quad k = 1, \dots, N \\ & y_k \geq 0, \quad k = 1, \dots, N \\ & z_k \geq 0, \quad k = 1, \dots, N \\ & x \in \mathcal{B}. \end{aligned} \right\} \quad (2.179)$$

The proof of the equivalence of (2.177) and (2.179) runs along the same lines as the proof for the equivalence of (2.173) and (2.174). For the equivalence of (2.176) and (2.178) it is sufficient to remark, that for any feasible solution  $(\hat{x}, \hat{y}_k, \hat{z}_k, k = 1, \dots, N)$  of (2.178),  $\hat{x}$  is feasible in (2.176), due to the following inequality:

$$\sum_{k=1}^N p_k |t^k x - h^k| = \sum_{k=1}^N p_k |y_k - z_k| \leq \sum_{k=1}^N p_k (y_k + z_k) \leq \kappa.$$

Analogously to the quadratic measure, we consider the variant which measures absolute deviations from the expected value and is called *mean absolute deviation* (MAD):

$$\rho_{\text{MAD}}(\vartheta) := \mathbb{E}[|\vartheta - \mathbb{E}[\vartheta]|], \quad \vartheta \in \mathcal{L}_1^1.$$

The evaluation function becomes

$$A_d(x) = |(\eta - \mu)^T x - (\xi - \mu_{n+1})|$$

leading to convex optimization problems, which are analogous to (2.172) and (2.173). The linear programming formulations for the case of a finite discrete distribution coincide with (2.178) and (2.179) when we set  $t^k := (\hat{\eta}^k - \mu)^T$  and  $h^k := \hat{\xi}^k - \mu_{n+1}$ .

An important special case in practice (for instance, in portfolio optimization in finance) is the case  $\xi \equiv 0$  thus leading to the deviation measure

$$A_d(x) = |(\eta - \mu)^T x| = |\eta^T x - \mu^T x|. \quad (2.180)$$

For the discretely distributed case we formulate the particular form of the optimization problems explicitly. Let  $t^k := \hat{\eta}^k$  (note that in (2.178) and (2.179) we have had  $t^k = (\hat{\eta}^k)^T$ ). (2.176) and (2.177) have the form now

$$\left. \begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & \sum_{k=1}^N p_k |(t^k - \mu)^T x| \leq \kappa \\ & x \in \mathcal{B} \end{aligned} \right\} \quad (2.181)$$

and

$$\left. \begin{array}{ll} \min & \sum_{k=1}^N p_k |(t^k - \mu)^T x| \\ \text{s.t.} & x \in \mathcal{B}, \end{array} \right\} \quad (2.182)$$

respectively. The equivalent linear programming formulations can easily be obtained from (2.178) and (2.179), by substituting  $t^k x$  with  $(t^k - \mu)^T x$  there.

Models of this type have first been proposed in the framework of portfolio optimization in finance by Konno and Yamazaki [194]. In this paper the authors propose a variant for the equivalent linear program (2.179) (with the substitution described above), by introducing fewer auxiliary variables on the cost of a larger amount of constraints, as follows:

$$\left. \begin{array}{ll} \min & \sum_{k=1}^N p_k y_k \\ \text{s.t.} & (t^k - \mu)^T x + y_k \geq 0, \quad k = 1, \dots, N \\ & (t^k - \mu)^T x - y_k \leq 0, \quad k = 1, \dots, N \\ & x \in \mathcal{B}. \end{array} \right\} \quad (2.183)$$

The equivalence with (2.179) can easily be seen, for instance, by considering separately the cases  $(t^k - \mu)^T x \geq 0$  and  $(t^k - \mu)^T x < 0$ .

Let us assume next that  $\eta$  has a non-degenerate multivariate normal distribution and let  $x \in \mathbb{R}^n$  be fixed. Then the random variable  $\eta^T x$ , being a linear transformation of a random vector with a non-degenerate normal distribution, is normally distributed (see Section 2.2.3). We obviously have  $\hat{\mu} := \mathbb{E}[\eta^T x] = \mu^T x$  and  $\hat{\sigma}^2 := \text{Var}[\eta^T x] = x^T V x$ . An easy computation gives:

$$\begin{aligned} \mathbb{E}[|\eta^T x - \mu^T x|] &= \hat{\sigma} \mathbb{E} \left[ \left| \frac{\eta^T x - \hat{\mu}^T x}{\hat{\sigma}} \right| \right] \\ &= \frac{\hat{\sigma}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |z| e^{-\frac{z^2}{2}} dz = \sqrt{\frac{2}{\pi}} \hat{\sigma} \int_0^{\infty} z e^{-\frac{z^2}{2}} dz \\ &= \sqrt{\frac{2}{\pi}} \hat{\sigma} = \sqrt{\frac{2}{\pi}} \sqrt{x^T V x}. \end{aligned}$$

This implies that for a non-degenerate normal distribution the models with absolute deviation and those with quadratic deviation are equivalent. Note, however, that due to the scaling factor  $\sqrt{2\pi}$  above, in the model (2.168) with a quadratic constraint, a scaling in the parameter  $\kappa$  has to be accounted for.

From the statistical point of view the natural measure for absolute deviations would be the absolute deviation from the median, instead of the expected value. The difficulty is that we are dealing with linear combinations of random variables  $\eta^T x = \sum_{i=1}^n \eta_i x_i$ . The median of  $\eta^T x$  is in general by no means equal to the linear com-



bination of the medians of the components of  $\eta$ . This makes it extremely difficult to build numerically tractable median-based optimization problems of the deviation type.

### 2.5.3 Quadratic semi-deviation

In both of the previous sections we employed risk measures which penalized deviations in both directions. The quadratic risk measure  $Q(x) = \sqrt{\mathbb{E}[(\eta^T x - \xi)^2]}$  (2.157) evaluates upper- and lower deviations of  $\eta^T x$  with respect to the target random variable  $\xi$  in the same manner. This observation holds also for the standard deviation  $\sigma(x) = \sqrt{\mathbb{E}[(\eta^T x - \mu^T x)^2]}$  (2.167) with respect to the deterministic target  $\mu^T x$ , and for the absolute-deviation counterparts  $A(x)$  (2.171), and  $A_d(x)$  (2.180). All of these risk measures model risk as deviation from a target, irrespectively of the direction of this deviation.

In many modeling situations, however, the direction of deviation matters. In such cases one of them is favorable (gain) and the other is disadvantageous (loss).

We introduce the following risk measures for random variables:

$$\begin{aligned}\rho_Q^+(\vartheta) &:= \sqrt{\mathbb{E}[(\vartheta^+)^2]}, \\ \rho_Q^-(\vartheta) &:= \sqrt{\mathbb{E}[(\vartheta^-)^2]}, \quad \vartheta \in \mathcal{L}_1^2,\end{aligned}$$

both of them being defined on the linear space of random variables with finite variance and with  $z^- = \max\{0, -z\}$ ,  $z^+ = \max\{0, z\}$  standing for the negative- and positive part for a real number  $z$ , respectively.

Let us assume that the second moments for  $(\eta^T, \xi)$  exist. The corresponding evaluation functions, denoted by  $Q^-(x)$  and  $Q^+(x)$ , respectively, are defined as

$$\begin{aligned}Q^+(x) &:= \sqrt{\mathbb{E}[(\eta^T x - \xi)^+]^2} \\ Q^-(x) &:= \sqrt{\mathbb{E}[(\eta^T x - \xi)^-]^2}.\end{aligned}\tag{2.184}$$

These measures are interpreted as measuring the upper/lower deviation between  $\eta^T x$  and  $\xi$ . Both  $Q^+(x)$  and  $Q^-(x)$  are convex functions; this will be proved in a general framework in Section 2.7.2, see Propositions 2.47. and 2.50. there.

Let us assume that negative values of the random variable  $\zeta(x, \eta, \xi) := \eta^T x - \xi$  represent losses. Then the following prototype optimization problems result

$$\left. \begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & \sqrt{\mathbb{E}[(\eta^T x - \xi)^-]^2} \leq \sqrt{\kappa} \\ & x \in \mathcal{B} \end{aligned} \right\} \tag{2.185}$$

and

$$\left. \begin{array}{ll} \min & \sqrt{\mathbb{E}[(\eta^T x - \xi)^-]^2} \\ \text{s.t.} & x \in \mathcal{B}, \end{array} \right\} \quad (2.186)$$

both of which are convex optimization problems. They can be equivalently written, due to the fact that the function  $\sqrt{z}$  is strictly monotonically increasing, as

$$\left. \begin{array}{ll} \min & c^T x \\ \text{s.t.} & \mathbb{E}[(\eta^T x - \xi)^-]^2 \leq \kappa \\ & x \in \mathcal{B} \end{array} \right\} \quad (2.187)$$

and

$$\left. \begin{array}{ll} \min & \mathbb{E}[(\eta^T x - \xi)^-]^2 \\ \text{s.t.} & x \in \mathcal{B}. \end{array} \right\} \quad (2.188)$$

Let us discuss the case when  $(\eta, \xi)$  has a discrete distribution specified in (2.175) on page 164. In this case our problems assume the form:

$$\left. \begin{array}{ll} \min & c^T x \\ \text{s.t.} & \sum_{k=1}^N p_k ((t^k x - h^k)^-)^2 \leq \kappa, \quad k = 1, \dots, N \\ & x \in \mathcal{B} \end{array} \right\} \quad (2.189)$$

and

$$\left. \begin{array}{ll} \min & \sum_{k=1}^N p_k ((t^k x - h^k)^-)^2 \\ \text{s.t.} & x \in \mathcal{B}. \end{array} \right\} \quad (2.190)$$

By introducing auxiliary variables  $y_k$ ,  $k = 1, \dots, N$ , these problems can be written equivalently as follows:

$$\left. \begin{array}{ll} \min & c^T x \\ \text{s.t.} & \sum_{k=1}^N p_k y_k^2 \leq \kappa, \quad k = 1, \dots, N \\ & t^k x + y_k \geq h^k, \quad k = 1, \dots, N \\ & y_k \geq 0, \quad k = 1, \dots, N \\ & x \in \mathcal{B} \end{array} \right\} \quad (2.191)$$

and

$$\left. \begin{array}{ll} \min & \sum_{k=1}^N p_k y_k^2 \\ \text{s.t.} & t^k x + y_k \geq h^k, \quad k = 1, \dots, N \\ & y_k \geq 0, \quad k = 1, \dots, N \\ & x \in \mathcal{B}. \end{array} \right\} \quad (2.192)$$

Problem (2.192) is a convex quadratic programming problem whereas (2.191) is a quadratically constrained convex optimization problem.

For proving the equivalence of (2.189) and (2.191), as well as of (2.190) and (2.192) let us make the following observation. If  $(\hat{x}, \hat{y}_k, k = 1, \dots, N)$  is a feasible solution of either (2.191) or (2.192), then the constraints imply the inequality  $(t^k \hat{x} - h^k)^- \leq y_k$  for all  $k$ . Consequently, in both cases

$$\sum_{k=1}^N p_k ((t^k \hat{x} - h^k)^-)^2 \leq \sum_{k=1}^N p_k \hat{y}_k^2$$

holds. From this the equivalence follows in a straightforward way; the detailed proof is left as an easy exercise to the reader.

Analogously as in both previous sections we discuss the variants measuring deviations from the expected value:

$$\begin{aligned} \rho_{\text{Std}}^+(\vartheta) &:= \sigma^+(x) := \sqrt{\mathbb{E}[(\vartheta - \mathbb{E}[\vartheta])^+]^2}, \\ \rho_{\text{Std}}^-(\vartheta) &:= \sigma^-(x) := \sqrt{\mathbb{E}[(\vartheta - \mathbb{E}[\vartheta])^-]^2}, \quad \vartheta \in \mathcal{L}_1^2. \end{aligned}$$

These are called *upper standard semi-deviation* and *lower standard semi-deviation*, respectively. The evaluation functions are obtained by performing the substitution  $(\eta - \mu)^T x - (\xi - \mu_{n+1})$  for  $\vartheta$ , which leads to convex optimization problems analogous to (2.187) and (2.173). In the case of a finite discrete distribution, the linear programming formulations coincide with (2.191) and (2.190) provided that the definitions  $t^k := (\hat{\eta}^k - \mu)^T$  and  $h^k := \hat{\xi}^k - \mu_{n+1}$  are used.

We discuss the important special case where  $\xi \equiv 0$  holds separately. The valuation functions take the form:

$$\begin{aligned} \sigma^+(x) &:= \sqrt{\mathbb{E}[(\eta^T x - \mu^T x)^+]^2} \\ \sigma^-(x) &:= \sqrt{\mathbb{E}[(\eta^T x - \mu^T x)^-]^2}, \end{aligned} \tag{2.193}$$

which are interpreted as measuring the upper/lower deviation between  $\eta^T x$  and its expected value  $\mu^T x$ . Because of its importance in practice we formulate the optimization problems for the case when  $\xi$  has a finite discrete distribution explicitly. With  $t^k$  now considered as a column vector, (2.189) and (2.190) have the form:

$$\left. \begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & \sum_{k=1}^N p_k ((t^k - \mu)^T x)^-)^2 \leq \kappa, \quad k = 1, \dots, N \\ & x \in \mathcal{B} \end{aligned} \right\} \tag{2.194}$$

and

$$\left. \begin{aligned} \min \quad & \sum_{k=1}^N p_k ((t^k - \mu)^T x)^-)^2 \\ \text{s.t.} \quad & x \in \mathcal{B} \end{aligned} \right\} \tag{2.195}$$

whereas (2.191) and (2.192) assume the form

$$\left. \begin{array}{l} \min c^T x \\ \text{s.t.} \quad \sum_{k=1}^N p_k y_k^2 \leq \kappa, \quad k = 1, \dots, N \\ (t^k - \mu)^T x + y_k \geq 0, \quad k = 1, \dots, N \\ y_k \geq 0, \quad k = 1, \dots, N \\ x \in \mathcal{B} \end{array} \right\} \quad (2.196)$$

and

$$\left. \begin{array}{l} \min \sum_{k=1}^N p_k y_k^2 \\ \text{s.t.} \quad (t^k - \mu)^T x + y_k \geq 0, \quad k = 1, \dots, N \\ y_k \geq 0, \quad k = 1, \dots, N \\ x \in \mathcal{B}. \end{array} \right\} \quad (2.197)$$

The importance of introducing this type of risk measures has first been recognized by Markowitz [217] who also applied them in financial portfolio optimization.

### 2.5.4 Absolute semi-deviation

Similarly to the way for constructing a semi-deviation variant for quadratic deviation, we get the following semi-deviation measures:

$$\begin{aligned} \rho_{\text{sic}}^+(\vartheta) &:= \mathbb{E}[\vartheta^+], \\ \rho_{\text{sic}}^-(\vartheta) &:= \mathbb{E}[\vartheta^-], \quad \vartheta \in \mathcal{L}_1^1, \end{aligned}$$

defined on the space of random variables with finite expected value and with  $z^- = \max\{0, -z\}$ ,  $z^+ = \max\{0, z\}$  for any real number  $z$ . Note that we do not obtain new risk measures: These risk measures have been already discussed in Section 2.4.1, in connection with integrated chance constraints, see (2.114) on page 143. Now we consider them again, this time in relation with the absolute-deviation risk measure  $\rho_A$  defined in (2.170). Using the relations  $z = z^+ - z^-$  and  $|z| = z^+ + z^-$  we obtain that  $z^- = \frac{1}{2}(|z| - z)$  and  $z^+ = \frac{1}{2}(|z| + z)$  hold. Thus we have

$$\begin{aligned} \rho_{\text{sic}}^+(\vartheta) &= \frac{1}{2} (\mathbb{E}[|\vartheta|] + \mathbb{E}[\vartheta]) = \frac{1}{2} (\rho_A(\vartheta) + \mathbb{E}[\vartheta]) \\ \rho_{\text{sic}}^-(\vartheta) &= \frac{1}{2} (\mathbb{E}[|\vartheta|] - \mathbb{E}[\vartheta]) = \frac{1}{2} (\rho_A(\vartheta) - \mathbb{E}[\vartheta]). \end{aligned} \quad (2.198)$$

The evaluation function  $A(x)$  (2.171) defined on page 163 has now the semi-deviation counterparts:

$$\begin{aligned}
K(x) &= \mathbb{E}[(\eta^T x - \xi)^+] = \frac{1}{2} \left( \mathbb{E}[|\eta^T x - \xi|] + (\mu^T x - \mu_{n+1}) \right) \\
H(x) &= \mathbb{E}[(\eta^T x - \xi)^-] = \frac{1}{2} \left( \mathbb{E}[|\eta^T x - \xi|] - (\mu^T x - \mu_{n+1}) \right),
\end{aligned} \tag{2.199}$$

which are the separate integrated probability functions defined in Section 2.4.1 on page 143. The following relations hold:

$$\begin{aligned}
K(x) &= \frac{1}{2} \left( A(x) + (\mu^T x - \mu_{n+1}) \right) \\
H(x) &= \frac{1}{2} \left( A(x) - (\mu^T x - \mu_{n+1}) \right).
\end{aligned} \tag{2.200}$$

According to Proposition 2.45. on page 163  $A(\cdot)$  is a convex function, consequently both  $K(x)$  and  $H(x)$  are convex functions, too.

Turning our attention to the case when the lower/upper absolute deviation is measured with respect to the expected value, we obviously have (see (2.198))

$$\begin{aligned}
\rho_{\text{MAD}}^+(\vartheta) &:= \mathbb{E}[(\vartheta - \mathbb{E}[\vartheta])^+] = \frac{1}{2} \rho_{\text{MAD}}(\vartheta) \\
\rho_{\text{MAD}}^-(\vartheta) &:= \mathbb{E}[(\vartheta - \mathbb{E}[\vartheta])^-] = \frac{1}{2} \rho_{\text{MAD}}(\vartheta).
\end{aligned}$$

This implies that the optimization model for minimizing the corresponding evaluation function will deliver the same results as its mean–absolute–deviation counterpart. With the valuation function in the constraint, the only difference with respect to (2.181) will be the right–hand–side of this constraint: with the semi–deviation measure this will be  $2\kappa$ .

## Exercises

**2.10.** Give a direct, detailed proof for the fact that  $A(x) := \mathbb{E}[|\eta^T x - \xi|]$  is a convex function.

**2.11.** Complete the proof of the equivalence of (2.189) and (2.191), as well as of (2.190) and (2.192). The main ingredient of the proof is sketched in the paragraph next to (2.192).

## 2.6 Modeling risk and opportunity

The different SLP model classes in the previous sections have been identified as follows: a quality measure  $\rho$  has been chosen first which characterizes the model class. Based on the selected quality measure, the corresponding evaluation function  $V(x) := \rho(\zeta(x, \xi))$  was utilized in building SLP models belonging to the class of models.

In this section we will take a look on some modeling issues concerning SLP models. For the sake of simplicity we will consider the following pair of prototype

problems

$$\begin{cases} \min V(x) \\ \text{s.t.} & x \in \mathcal{B} \end{cases} \quad \begin{cases} \max V(x) \\ \text{s.t.} & x \in \mathcal{B} \end{cases} \quad (2.201)$$

with the evaluation functions  $V$  in the objective. Analogous reasoning applies for SLP models involving constraints with  $V$ .

Before proceeding let us emphasize that the two problems in (2.201) are substantially different from the numerical point of view. Assuming, for instance, that  $V$  is a nonlinear convex function, this implies that the minimization problem is in general much more easier to solve numerically than its maximization counterpart. Applying the usual trick for transforming the maximization problem into a minimization problem involving  $-V$  in the objective, does not help in this respect, of course.

Let us point out next that, from the modeling viewpoint, the mere definition and mathematical properties of a quality measure  $\rho$  do not a priori imply a selection between the two possible models in (2.201). To see this, consider the standard deviation  $\rho_{\text{Std}}(\vartheta) := \sigma(\vartheta) := \mathbb{E}[(\vartheta - \mathbb{E}[\vartheta])^2]^{\frac{1}{2}}$  as a quality measure, discussed in Section 2.5.1. Notice that the implied evaluation function  $V$  is a convex function. With this evaluation function, the SLP model (2.166) on page 162 corresponds to the minimization formulation in (2.201). The usage of this model presupposes the following modeling attitude: the modeler interprets any deviation from the expected value as risk, quantifies the deviations by choosing the standard deviation as quality measure, and seeks to minimize this quality measure. In this modeling context, the quality measure  $\rho_{\text{Std}}$  can be interpreted as a risk measure. Note that assuming a symmetric distribution, a large standard deviation indicates that  $\vartheta$  exhibits large deviations both in the upward and downward direction with respect to the expected value. Consider now a gambler. For she/he the upward deviations represent an opportunity for winning, therefore larger standard deviations will be preferred to smaller values. This modeler would choose the maximization problem in (2.201). Consequently, for such a modeler the interpretation of the same quality measure  $\rho_{\text{Std}}$  is clearly an opportunity measure. The modeler faces a non-convex optimization problem.

In the previous example the same quality measure served simultaneously as risk- and opportunity-measure, the sole difference was the way, how it has been used for building SLP models. Both for the risk-averse modeler and for the gambler the standard deviation is not the best way for building an SLP model. To see this, and to further explore the ways for modeling risk and opportunity, let us assume that

- negative values of  $\zeta(x, \xi) - \mathbb{E}[\zeta(x, \xi)]$  are interpreted as something unpleasant, like costs, loss in wealth, or loss in health;
- positive values of  $\zeta(x, \xi) - \mathbb{E}[\zeta(x, \xi)]$  quantify something desirable, like monetary gains or stability of an engineering structure;
- $\zeta(x, \xi) - \mathbb{E}[\zeta(x, \xi)] = 0$  expresses neutrality in the risk-opportunity aspect.

Instead of the standard deviation, in this situation it makes sense to choose the lower- and upper standard semi-deviations (see Section 2.5.3) as quality measures. The risk-averse modeler would choose the lower semi-deviation  $\rho_Q^-(\vartheta) := \mathbb{E}[(\vartheta^-)^2]^{\frac{1}{2}}$ , interpreted as a risk measure. The corresponding optimization problem is the minimization problem in (2.201). A modeler who does not care

for risk would choose the upper semi-deviation  $\rho_Q^+(\vartheta) := \mathbb{E}[(\vartheta^+)^2]^{\frac{1}{2}}$  with the corresponding maximization problem in (2.201). The corresponding evaluation functions are convex functions for both the lower- and for the upper semi-deviation, see the discussion on page 167. Therefore, again, the risk-averse modeler faces a convex optimization problem whereas the modeler neglecting risk has a non-convex optimization problem to solve. The idea of combining the two quality measures, for instance as  $\rho_Q^- - \lambda \rho_Q^+$  with  $\lambda > 0$ , and minimizing the resulting evaluation function, still results in a non-convex optimization problem.

Another possibility for employing a suitable quality measure is to work with separate integrated probability functions, see Section 2.4.1. The risk-averse modeler would choose  $\rho_{\text{sic}}^-(\vartheta) := \mathbb{E}[\vartheta^-]$  with the corresponding minimization problem whereas her/his risk-seeking counterpart would employ  $\rho_{\text{sic}}^+(\vartheta) := \mathbb{E}[\vartheta^+]$  and the maximization problem. Both corresponding evaluation functions are convex, see Proposition 2.39., therefore analogous comments apply as for the semi-deviations. There is, however, an essential difference: now it makes sense to combine the two quality measures. This leads to the quality measure  $\rho_{\text{sic}}^\alpha$  discussed on page 146 with  $\alpha \in [0, 1]$ . For  $\alpha \in [0, \frac{1}{2}]$  it serves as a risk measure with a convex evaluation function whereas for  $\alpha \in (\frac{1}{2}, 1]$  it can be interpreted as quantifying opportunity with a corresponding concave evaluation function.

Finally let us discuss the usage of probability functions in modeling. Concerning separate probability functions, we have seen in Section 2.2.3, that, for certain special cases convex programming problems arise. This is true both for the risk-averse and for the risk-seeking attitude. For joint probability constraints the situation is different, see Section 2.2.5. Convex programming problems can only be obtained when interpreting the quality measure as a measure of opportunity, that means, the evaluation function is to be maximized.

## 2.7 Risk measures

We consider random variables of the form

$$\zeta(x, \eta, \xi) := \eta^T x - \xi,$$

where  $(\eta^T, \xi)$  is an  $n+1$ -dimensional random vector defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ ;  $\eta$  denotes an  $n$ -dimensional random vector and  $\xi$  is a random variable. Whenever the expected value of  $(\eta^T, \xi)$  exists we will employ the notation  $\mu := \mathbb{E}[\eta] \in \mathbb{R}^n$  and  $\mu_{n+1} := \mathbb{E}[\xi] \in \mathbb{R}$ .

In the previous sections we used a two-step scheme in presenting the various stochastic programming model classes. In a first step we have specified a function  $\rho : Y \rightarrow \mathbb{R}$  for evaluating random variables with  $Y$  being some linear space of random variables, defined on a probability space  $(\Omega, \mathcal{F}, P)$ . We have called  $\rho$  a quality measure concerning random variables. In a second step, provided that  $\zeta(x, \eta, \xi) \in Y$  holds for all  $x$ , we have substituted  $\zeta(x, \eta, \xi)$  into  $\rho$  thus getting the evaluation func-

tion  $V$ ,  $V(x) := \rho(\zeta(x, \eta, \xi))$ .  $V$  has been subsequently used for building SLP models. Assuming that  $\rho$  quantifies risk,  $V$  has been built into SLP models as follows: If in the objective, then  $V(x)$  was minimized and if in a constraint then constraints of the type  $V(x) \leq \kappa$  were employed. This modeling attitude justifies the usage of the term “risk measure” for  $\rho$ . For optimization models involving  $V$  in the above outlined fashion, the (generalized) convexity of  $V$  is clearly an advantageous property. It leads to optimization problems for which we have good chances for finding an efficient numerical solution procedure.

For fixed  $(\eta, \xi)$ ,  $\zeta(\cdot, \eta, \xi)$  is a linear-affine function, thus there is a close relation between structural properties of  $\rho$  and (generalized) convexity properties of  $V$ . The purpose of this section is to discuss properties of various risk measures and their impact on the evaluation function.

Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\vartheta$  be a random variable on it. The distribution function of  $\vartheta$  will be denoted by  $F_\vartheta$  and  $\Theta$  denotes the support of  $\vartheta$ . Recall, that  $Y$  has been chosen as one of the linear spaces listed in (2.6) on page 73.

A function  $g : X \rightarrow \mathbb{R}$ , defined on a linear space  $X$ , is called *positively homogeneous*, if for any  $\lambda \geq 0$  and  $x \in X$ , the relation  $g(\lambda x) = \lambda g(x)$  holds.  $g$  is called *subadditive*, if for any  $x, y \in X$  the inequality  $g(x + y) \leq g(x) + g(y)$  holds. For later reference the following simple facts are formulated as an assertion:

**Proposition 2.46.** *Let  $g : X \rightarrow \mathbb{R}$  be a function defined on a linear space  $X$ . Then*

- a) if  $g$  is both positively homogeneous and subadditive then it is convex.*
- b) Suppose that  $g$  is positively homogeneous and convex. This implies subadditivity.*

*Proof:* In fact, let  $x, y \in X$  and  $\lambda \in (0, 1)$  then we have

$$g(\lambda x + (1 - \lambda)y) \leq g(\lambda x) + g((1 - \lambda)y) = \lambda g(x) + (1 - \lambda)g(y),$$

where the inequality follows from subadditivity and the equality from positive homogeneity. This proves a). Suppose that  $g$  is positively homogeneous and convex and let  $x, y \in X$  then

$$g(x + y) = g(2[\frac{1}{2}x + \frac{1}{2}y]) = 2g(\frac{1}{2}x + \frac{1}{2}y) \leq g(x) + g(y)$$

from which b) follows. □

The next proposition establishes a relation between properties of  $\rho$  and properties of the corresponding evaluation function  $V$ .

**Proposition 2.47.** *Let  $Y$  be a linear space of random variables and  $\rho : Y \rightarrow \mathbb{R}$  a real-valued function on  $Y$ . Assume that  $\eta^T x - \xi \in Y$  holds for all  $x$  and let  $V(x) := \rho(\eta^T x - \xi)$ . Then we have:*

- a) If  $\rho$  is convex then  $V$  is convex too.*
- b) If  $\xi \equiv 0$  and  $\rho$  is subadditive then  $V$  is also subadditive.*
- c) If  $\xi \equiv 0$  and  $\rho$  is positively homogeneous then  $V$  is also positively homogeneous.*



*Proof:*

a) Let  $x, y \in \mathbb{R}^n$  and  $\lambda \in [0, 1]$  then we have

$$\begin{aligned} V(\lambda x + (1 - \lambda)y) &= \rho(\eta^T(\lambda x + (1 - \lambda)y) - \xi) \\ &= \rho(\lambda(\eta^T x + \xi) + (1 - \lambda)(\eta^T y + \xi)) \\ &\leq \lambda \rho(\eta^T x + \xi) + (1 - \lambda) \rho(\eta^T y + \xi) \\ &= \lambda V(x) + (1 - \lambda)V(y). \end{aligned}$$

Assertions b) and c) follow similarly.  $\square$

Notice that in the above assertions the stated properties of  $V$  hold for any probability distribution of  $\vartheta \in \mathcal{Y}$ . Thus we obtain convex SLP problems under the sole assumption  $\vartheta \in \mathcal{Y}$ . By proving the convexity of a specific risk measure  $\rho$ , we obtain alternative proofs of convexity of the corresponding SLP problems discussed in the previous sections.

### 2.7.1 Risk measures in finance

In financial theory and praxis, more closely in portfolio optimization, an increasing effort in research is devoted to identify those properties of risk measures, which are distinguishing features. The general aim of the research is twofold. On the one hand, the goal is to develop an axiomatically founded risk theory in finance. On the other hand, the aim is to provide guidelines for practitioners for choosing an appropriate risk measure in their daily work and to support the construction of appropriate standards for risk management in the finance industry. Several different definitions and systems of axioms have been proposed in the financial literature. Below we simply list some of the current definitions without discussing their intuitive background and implications, these being application-specific. We assume throughout that positive values of the random variables  $\vartheta \in \mathcal{Y}$  represent losses.

Kijima and Ohnisi [186] propose the following definition:  $\rho$  is a risk measure, if the following properties hold for any  $\vartheta, \vartheta_1, \vartheta_2 \in \mathcal{Y}$  and  $\lambda, C \in \mathbb{R}$ ,  $\lambda \geq 0$ :

$$\begin{aligned} (K1) \quad & \rho(\vartheta_1 + \vartheta_2) \leq \rho(\vartheta_1) + \rho(\vartheta_2) \quad (\text{subadditivity}) \\ (K2) \quad & \rho(\lambda \vartheta) = \lambda \rho(\vartheta) \quad (\text{positive homogeneity}) \\ (K3) \quad & \rho(\vartheta) \geq 0 \quad (\text{nonnegativity}) \\ (K4) \quad & \rho(\vartheta + C) = \rho(\vartheta) \text{ for } C \geq 0 \quad (\text{shift invariance}) \end{aligned} \tag{2.202}$$

The important issue of axiomatic foundation of risk measures has first been addressed in the seminal paper of Artzner, Delbaen, Eber, and Heath [7]. The authors propose the axioms below and explore their implications. We formulate the axioms for random variables representing losses, whereas in the original paper the interpretation is future value.

$$\begin{aligned}
(A1) \quad & \rho(\vartheta_1 + \vartheta_2) \leq \rho(\vartheta_1) + \rho(\vartheta_2) \quad (\text{subadditivity}) \\
(A2) \quad & \rho(\lambda \vartheta) = \lambda \rho(\vartheta) \quad (\text{positive homogeneity}) \\
(A3) \quad & \text{If } \vartheta_1 \leq \vartheta_2 \text{ then } \rho(\vartheta_1) \leq \rho(\vartheta_2) \quad (\text{monotonicity}) \\
(A4) \quad & \rho(\vartheta + C) = \rho(\vartheta) + C \quad (\text{translation invariance})
\end{aligned} \tag{2.203}$$

The authors call a function  $\rho$  for which the above axioms hold a *coherent risk measure*. Concerning SLP models in general, in an intuitive sense the axiom A4 looks rather unusual. The reason for including it in this form is that the authors consider capital requirement problems, see [7]. For distinguishing between the different requirements concerning translation in (2.202) and (2.203), we use the terms “shift invariance” and “translation invariance”, respectively.

In the system of axioms of Föllmer and Schied [94], [95], subadditivity and positive homogeneity is replaced by the weaker requirement of convexity

$$\begin{aligned}
(F1) \quad & \rho(\vartheta) \text{ is a convex function} \quad (\text{convexity}) \\
(F2) \quad & \text{If } \vartheta_1 \leq \vartheta_2 \text{ then } \rho(\vartheta_1) \leq \rho(\vartheta_2) \quad (\text{monotonicity}) \\
(F3) \quad & \rho(\vartheta + C) = \rho(\vartheta) + C \quad (\text{translation invariance})
\end{aligned} \tag{2.204}$$

leading to *convex risk measures*. Coherent risk measures are obviously convex; a convex risk measure is coherent, if it is positively homogeneous (see Proposition 2.46.).

Rockafellar, Uryasev, and Zabarankin [284] introduce the notion of *deviation measure* for  $\mathcal{R} = \mathcal{L}_1^1$ . Their axioms are

$$\begin{aligned}
(D1) \quad & \rho(\vartheta_1 + \vartheta_2) \leq \rho(\vartheta_1) + \rho(\vartheta_2) \quad (\text{subadditivity}) \\
(D2) \quad & \rho(\lambda \vartheta) = \lambda \rho(\vartheta) \quad (\text{positive homogeneity}) \\
(D3) \quad & \rho(\vartheta) > 0 \text{ for } \vartheta \text{ non-constant,} \\
& \rho(\vartheta) = 0 \text{ otherwise} \quad (\text{nonnegativity}) \\
(D4) \quad & \rho(\vartheta + C) = \rho(\vartheta) \quad (\text{shift invariance})
\end{aligned} \tag{2.205}$$

The authors also define an associated risk measure called *expectation-bounded risk measure*, see [284], and explore the implications in portfolio theory. Notice that for a deviation measure the axioms (2.202) hold; the axioms for a deviation measure can be considered as a refinement (restriction) of (2.202).

Due to the different prescription for the case of a translation, the set of risk measures obeying (2.202) or (2.205) and the risk measures for which either (2.203) or (2.204) hold, are disjunct sets.

We feel that there is not much chance that a general definition of a risk measure can be given, which would be acceptable also beyond the field of finance. From our general stochastic programming point of view, the convexity of a risk-measure is surely a desirable property. Proposition 2.47. implies, namely, that the SLP models, which are built on the basis of such a measure, are convex optimization problems. From this viewpoint, a risk measure can be considered as more valuable, when beyond serving as a diagnostic metric, it can also be built into efficiently solvable optimization models which involve, for instance, minimizing risk. Without exception, all of the above definitions correspond to risk measures of this type.

### 2.7.2 Properties of risk measures

This section is devoted to discussing convexity properties of risk measures which have been utilized for building stochastic programming models. Unless explicitly referring to the axioms (2.204) of convex risk measures, under convexity we will simply mean convexity of the risk–measure–function  $\rho$ . We will use the following notation: the functions  $\varphi^+ : \mathbb{R} \rightarrow \mathbb{R}_+$  and  $\varphi^- : \mathbb{R} \rightarrow \mathbb{R}_+$  are defined as  $\varphi^+(z) := z^+$  and  $\varphi^-(z) := z^-$ , respectively, where  $z^+ = \max\{0, z\}$  and  $z^- = \max\{0, -z\}$  are the positive– and negative part of the real number  $z$ . Let further  $\varphi^A$  denote the absolute–value function  $\varphi^A(z) := |z|$  for all  $z \in \mathbb{R}$ . The relation  $\varphi^A = \varphi^+ + \varphi^-$  obviously holds. Note that  $\varphi^+$ ,  $\varphi^-$ , and  $\varphi^A$  are positively homogeneous and subadditive functions, therefore they are convex.

**Proposition 2.48.** *The following risk measures are positively homogeneous and subadditive. Moreover, they are also monotonously increasing and translation invariant. Consequently, they are convex risk measures in the sense of axioms (2.204) and being positively homogeneous they are also coherent according to axioms (2.203).*

- (A)  $\rho_E(\vartheta) := \mathbb{E}[\vartheta]$ ,  $\vartheta \in \mathcal{L}_1^1$ , (Section 2.4);
- (B)  $\rho_{\text{fat}}(\vartheta) := \max_{\bar{\vartheta} \in \Theta} \hat{\vartheta}$ ,  $\vartheta \in \mathcal{L}_1^\infty$  where  $\Theta$  is the support of  $\vartheta$  (Section 2.1, page 76);  
note that we have changed minimum to maximum for getting a risk measure.  
 $\rho_{\text{fat}}$  is called the maximum loss risk measure.
- (C)  $\rho_{\text{CVaR}}^\alpha(\vartheta) := v_c(\vartheta, \alpha) := \min_z [z + \frac{1}{1-\alpha} \mathbb{E}[(\vartheta - z)^+]]$ ,  $\vartheta \in \mathcal{L}_1^1$ ,  $0 < \alpha < 1$ ,  
(Section 2.4.3).

*Proof:*

(A): The assertion holds trivially because  $\rho_E$  is a linear function.

(B): Let  $\lambda \geq 0$  be a real number and  $\vartheta \in \mathcal{L}_1^\infty$  with support  $\Theta$ . Then  $\lambda\Theta$  is a closed set, therefore it is the support of  $\lambda\vartheta$ . For the definition of the operation  $\lambda\Theta$  see (2.89) on page 118. Thus we have

$$\rho_{\text{fat}}(\lambda\vartheta) = \max_{\bar{\vartheta} \in \lambda\Theta} \hat{\vartheta} = \max_{\bar{\vartheta} \in \Theta} \lambda \bar{\vartheta} = \lambda \rho_{\text{fat}}(\vartheta).$$

For proving subadditivity let  $\vartheta_1, \vartheta_2 \in \mathcal{L}_1^\infty$  with supports  $\Theta_1$  and  $\Theta_2$ , respectively. Let  $\Theta := \Theta_1 + \Theta_2$  where the sum of the two sets is defined according to (2.89) on page 118. From the discussion on that page it follows that  $\Theta$  is a closed set. Consequently,  $\Theta$  contains the support of  $\vartheta_1 + \vartheta_2$ . Thus we have

$$\begin{aligned} \rho_{\text{fat}}(\vartheta_1 + \vartheta_2) &= \max_{\hat{\vartheta}_1 + \hat{\vartheta}_2 \in \text{supp}\{\vartheta_1 + \vartheta_2\}} [\hat{\vartheta}_1 + \hat{\vartheta}_2] \\ &\leq \max_{\hat{\vartheta}_1 + \hat{\vartheta}_2 \in \Theta} [\hat{\vartheta}_1 + \hat{\vartheta}_2] \leq \max_{\bar{\vartheta}_1 \in \Theta_1} \bar{\vartheta}_1 + \max_{\bar{\vartheta}_2 \in \Theta_2} \bar{\vartheta}_2. \end{aligned}$$

For any real number  $C$ , the support of  $\vartheta + C$  is  $\Theta + \{C\}$ , from which the translation invariance immediately follows. If  $\vartheta_1(\omega) \leq \vartheta_2(\omega)$  holds for all  $\omega \in \Omega$  then we obviously have  $\rho_{\text{fat}}(\vartheta_1) \leq \rho_{\text{fat}}(\vartheta_2)$ .

(C): Let  $\lambda \geq 0$  and  $\vartheta \in \mathcal{L}_1$ . If  $\lambda = 0$ , then we have

$$\rho_{\text{CVaR}}^\alpha(\lambda \vartheta) = \rho_{\text{CVaR}}^\alpha(0) = \min_z [z + \frac{1}{1-\alpha} \mathbb{E}[(-z)^+]] = 0,$$

where the last equality follows from

$$z + \frac{1}{1-\alpha} \mathbb{E}[(-z)^+] = \begin{cases} z & \text{if } z \geq 0 \\ (1 - \frac{1}{1-\alpha})z & \text{if } z < 0. \end{cases}$$

Assuming  $\lambda > 0$  we have

$$\begin{aligned} \rho_{\text{CVaR}}^\alpha(\lambda \vartheta) &= \min_z [z + \frac{1}{1-\alpha} \mathbb{E}[(\lambda \vartheta - z)^+]] \\ &= \lambda \min_z [\frac{z}{\lambda} + \frac{1}{1-\alpha} \mathbb{E}[(\vartheta - \frac{z}{\lambda})^+]] \\ &= \lambda \min_y [y + \frac{1}{1-\alpha} \mathbb{E}[(\vartheta - y)^+]] = \lambda \rho_{\text{CVaR}}^\alpha(\vartheta). \end{aligned}$$

For proving subadditivity we utilize the fact (see Section 2.4.3) that the minimum in the definition is attained. Let  $\vartheta_1, \vartheta_2 \in \mathcal{L}_1$  and  $z_1, z_2$  be corresponding solutions of the minimization problem in the definition. For proving subadditivity it is sufficient to prove convexity (see Proposition 2.46.). Let  $0 < \lambda < 1$ ,  $z_\lambda$  be the minimum for  $\lambda \vartheta_1 + (1-\lambda)\vartheta_2$ , and  $\bar{z}_\lambda = \lambda z_1 + (1-\lambda)z_2$ . Utilizing the convexity of  $\varphi^+$  we get

$$\begin{aligned} \rho_{\text{CVaR}}^\alpha(\vartheta_1 + \vartheta_2) &= \min_z [z + \frac{1}{1-\alpha} \mathbb{E}[\varphi^+(\lambda \vartheta_1 + (1-\lambda)\vartheta_2 - z)]] \\ &= z_\lambda + \frac{1}{1-\alpha} \mathbb{E}[\varphi^+(\lambda \vartheta_1 + (1-\lambda)\vartheta_2 - z_\lambda)] \\ &\leq \bar{z}_\lambda + \frac{1}{1-\alpha} \mathbb{E}[\varphi^+(\lambda \vartheta_1 + (1-\lambda)\vartheta_2 - \bar{z}_\lambda)] \\ &\leq \rho_{\text{CVaR}}^\alpha(\vartheta_1) + \rho_{\text{CVaR}}^\alpha(\vartheta_2). \end{aligned}$$

Due to the fact that  $\varphi^+$  is a monotonically increasing function, the monotonicity of  $\rho_{\text{CVaR}}^\alpha$  follows immediately. Let  $C \in \mathbb{R}$  then we have

$$\begin{aligned} \rho_{\text{CVaR}}^\alpha(\vartheta + C) &= \min_z [z + \frac{1}{1-\alpha} \mathbb{E}[(\vartheta - (z - C))^+]] \\ &= C + \min_z [z - C + \frac{1}{1-\alpha} \mathbb{E}[(\vartheta - (z - C))^+]] \\ &= \rho_{\text{CVaR}}^\alpha(\vartheta) + C \end{aligned}$$

thus the translation invariance follows.  $\square$

For the next group of risk measures translation– or shift–invariance does not hold, but we have:

**Proposition 2.49.** *The risk measures listed below are positively homogeneous and subadditive and they are also monotonous.*

- (D1)  $\rho_{\text{sic}}^+(\vartheta) := \mathbb{E}[\vartheta^+]$ ,  $\vartheta \in \mathcal{L}_1^1$ , (Section 2.4.1);  
 (D2)  $\rho_{\text{sic}}^-(\vartheta) := \mathbb{E}[\vartheta^-]$ ,  $\vartheta \in \mathcal{L}_1^1$ , (Section 2.4.1);  
 (E)  $\rho_{\text{sic}}^\alpha(\vartheta) := \alpha\mathbb{E}[\vartheta] + (1-2\alpha)\mathbb{E}[\vartheta^+] = \alpha\rho_{\text{E}}(\vartheta) + (1-2\alpha)\rho_{\text{sic}}^+(\vartheta)$ ,  $\vartheta \in \mathcal{L}_1^1$ ,  
 $0 \leq \alpha \leq \frac{1}{2}$ , (Section 2.4.1).

*Proof:*

(D1): We have  $\rho_{\text{sic}}^+(\vartheta) = \mathbb{E}[\vartheta^+]$ . The function  $\vartheta^+$  being positively homogeneous and subadditive, as well as monotonously increasing, the assertion follows immediately. In fact, for proving subadditivity let  $\vartheta_1, \vartheta_2 \in \mathcal{L}_1$  then we have

$$\rho_{\text{sic}}^+(\vartheta_1 + \vartheta_2) = \mathbb{E}[(\vartheta_1 + \vartheta_2)^+] \leq \mathbb{E}[\vartheta_1^+ + \vartheta_2^+] = \rho_{\text{sic}}^+(\vartheta_1) + \rho_{\text{sic}}^+(\vartheta_2).$$

The proof for positive homogeneity is analogous.  $\rho_{\text{sic}}^+$  turns out to be a monotonically increasing function.

(D2): In this case  $\rho_{\text{sic}}^-(\vartheta) = \mathbb{E}[\vartheta^-]$  holds. The positive homogeneity and subadditivity of  $\vartheta^-$  implies these properties for  $\rho_{\text{sic}}^-$ .  $\vartheta^-$  being monotonously decreasing,  $\rho_{\text{sic}}^-$  is monotonically decreasing, too.

(E): This follows immediately from the linearity of the first term and from (D1).  $\square$

In the next group neither translation–invariance nor monotonicity holds. Nevertheless, we have

**Proposition 2.50.** *The following risk measures are positively homogeneous and subadditive.*

- (F)  $\rho_{\text{Q}}(\vartheta) := \sqrt{\mathbb{E}[\vartheta^2]}$ ,  $\vartheta \in \mathcal{L}_1^2$ , (Section 2.5.1);  
 (G)  $\rho_{\text{A}}(\vartheta) := \mathbb{E}[|\vartheta|]$ ,  $\vartheta \in \mathcal{L}_1^1$ , (Section 2.5.2);  
 (H)  $\rho_{\text{Q}}^+(\vartheta) := \sqrt{\mathbb{E}[(\vartheta^+)^2]}$  and  
 $\rho_{\text{Q}}^-(\vartheta) := \sqrt{\mathbb{E}[(\vartheta^-)^2]}$ ,  $\vartheta \in \mathcal{L}_1^1$ , (Section 2.5.3).

*Proof:* The positive homogeneity is trivial for all cases therefore we confine ourselves to proving subadditivity.

(F): Let  $\vartheta_1, \vartheta_2 \in \mathcal{L}_2$  then the Minkowski–inequality immediately yields

$$\begin{aligned} \rho_{\text{Q}}(\vartheta_1 + \vartheta_2) &= \left( \mathbb{E}[(\vartheta_1 + \vartheta_2)^2] \right)^{\frac{1}{2}} = \left( \mathbb{E}[|\vartheta_1 + \vartheta_2|^2] \right)^{\frac{1}{2}} \\ &\leq \left( \mathbb{E}[|\vartheta_1|^2] \right)^{\frac{1}{2}} + \left( \mathbb{E}[|\vartheta_2|^2] \right)^{\frac{1}{2}} = \rho_{\text{Q}}(\vartheta_1) + \rho_{\text{Q}}(\vartheta_2). \end{aligned}$$

(G): We have  $\rho_A(\vartheta) = \mathbb{E}[\varphi^A(\vartheta)]$  and the assertion follows from the subadditivity of  $\varphi^A$ .

(H): We prove the assertion for  $\rho_Q^+$ , the proof for  $\rho_Q^-$  is analogous. We utilize the subadditivity of  $\varphi^+$  and again the Minkowski-inequality:

$$\begin{aligned}\rho_Q^+(\vartheta_1 + \vartheta_2) &= (\mathbb{E}[(\vartheta_1 + \vartheta_2)^+]^2)^{\frac{1}{2}} \\ &\leq (\mathbb{E}[(\vartheta_1^+ + \vartheta_2^+)^2])^{\frac{1}{2}} \\ &\leq (\mathbb{E}[(\vartheta_1^+)^2])^{\frac{1}{2}} + (\mathbb{E}[(\vartheta_2^+)^2])^{\frac{1}{2}} = \rho_Q^+(\vartheta_1) + \rho_Q^+(\vartheta_2).\end{aligned}$$

□

Finally we turn our attention to the deviation measures in Section 2.5.

**Proposition 2.51.** *The following risk measures are deviation measures according to the axioms (2.205).*

- (I)  $\rho_{\text{Std}}(\vartheta) := \sigma(\vartheta) := \sqrt{\mathbb{E}[(\vartheta - \mathbb{E}[\vartheta])^2]}$ ,  $\vartheta \in \mathcal{L}_1^2$ ;
- (J)  $\rho_{\text{MAD}}(\vartheta) := \mathbb{E}[|\vartheta - \mathbb{E}[\vartheta]|]$ ,  $\vartheta \in \mathcal{L}_1^1$ ;
- (K)  $\rho_{\text{Std}}^+(\vartheta) := \sigma^+(x) := \sqrt{\mathbb{E}[(\vartheta - \mathbb{E}[\vartheta])^+]^2}$  and  $\rho_{\text{Std}}^-(\vartheta) := \sigma^-(x) := \sqrt{\mathbb{E}[(\vartheta - \mathbb{E}[\vartheta])^-]^2}$ ,  $\vartheta \in \mathcal{L}_1^1$ ;
- (L)  $\rho_{\text{MAD}}^+(\vartheta) := \mathbb{E}[(\vartheta - \mathbb{E}[\vartheta])^+] = \frac{1}{2}\rho_{\text{MAD}}(\vartheta)$  and  $\rho_{\text{MAD}}^-(\vartheta) := \mathbb{E}[(\vartheta - \mathbb{E}[\vartheta])^-] = \frac{1}{2}\rho_{\text{MAD}}(\vartheta)$ ,  $\vartheta \in \mathcal{L}_1^1$ .

*Proof:* Note that each one of these risk measures results from an already considered risk measure by substituting  $\vartheta$  by  $\vartheta - \mathbb{E}[\vartheta]$ . Therefore it is clear that each one is positively homogeneous and subadditive. All of them are nonnegative and can only be zero if  $\vartheta$  is constant. Finally the shift-property D4 holds trivially. □

Recall, that due to Proposition 2.46., the positive homogeneity and subadditivity of the risk functions considered so far implies that all of them are convex.

The risk measures listed below have been used for building SLP models but have not yet been considered:

- (M)  $\rho_P(\vartheta) := \mathbb{P}(\vartheta \geq 0)$ ,  $\vartheta \in \mathcal{V}$ , (Section 2.2);
- (N)  $\rho_{\text{cexp}}(\vartheta) := \mathbb{E}[-\vartheta \mid \vartheta < 0]$ ,  $\vartheta \in \mathcal{L}_1^1$ , (Section 2.4.2);
- (O)  $\rho_{\text{VaR}}^\alpha(\vartheta) := v(\vartheta, \alpha) := \min\{z \mid F_\vartheta(z) \geq \alpha\}$ ,  $\vartheta \in \mathcal{V}$ ,  $0 < \alpha < 1$ , (Section 2.3).

They are non-convex in general. Despite this fact, we have seen in the previous sections that under some assumptions concerning the probability distribution and parameter values, using these quality measures resulted in convex or in generalized convex optimization problems. The point is the following: having a convex risk measure  $\rho$ , this leads automatically to convex evaluation functions (see Proposition 2.47.) and thus to convex optimization problems. In other words, the convexity

of  $\rho$  is a sufficient condition for getting convex optimization problems. The convexity of  $\rho$  is by no means also necessary for this, as the convex optimization models, built on the basis of the above risk measures, and presented in the previous sections demonstrate.

At last let us consider risk measures for random vectors, introduced in Section 2.4.1 as

$$(P) \quad \rho_{\text{jic}}^+(\vartheta) := \mathbb{E}[\max_{1 \leq i \leq s} \vartheta_i^+] \text{ and } \rho_{\text{jic}}^-(\vartheta) := \mathbb{E}[\max_{1 \leq i \leq s} \vartheta_i^-], \quad \vartheta \in \mathcal{L}_1^1,$$

where  $\vartheta$  is now an  $s$ -dimensional random vector. These risk measures have the properties:

**Proposition 2.52.** *Both  $\rho_{\text{jic}}^+$  and  $\rho_{\text{jic}}^-$  are positively homogeneous and subadditive. Moreover, both of them are monotonous.*

*Proof:* The positive homogeneity is obvious. We prove the subadditivity for  $\rho_{\text{jic}}^+$ , the proof for  $\rho_{\text{jic}}^-$  is analogous. Let  $\vartheta^{(1)}, \vartheta^{(2)} \in \mathcal{L}_1^1$  then we have

$$\begin{aligned} \rho_{\text{jic}}^+(\vartheta^{(1)} + \vartheta^{(2)}) &= \mathbb{E}[\max_{1 \leq i \leq s} (\vartheta_i^{(1)} + \vartheta_i^{(2)})^+] \\ &\leq \mathbb{E} \left[ \max_{1 \leq i \leq s} [(\vartheta_i^{(1)})^+ + (\vartheta_i^{(2)})^+] \right] \\ &\leq \mathbb{E} \left[ \max_{1 \leq i \leq s} [(\vartheta_i^{(1)})^+] \right] + \mathbb{E} \left[ \max_{1 \leq i \leq s} [(\vartheta_i^{(2)})^+] \right] \\ &= \rho_{\text{jic}}^+(\vartheta^{(1)}) + \rho_{\text{jic}}^+(\vartheta^{(2)}) \end{aligned}$$

where for the first inequality we used the subadditivity of  $\varphi^+$  and the second inequality follows from the properties of the max operator. From the properties of  $\varphi^+$  and  $\varphi^-$  it is also clear that  $\rho_{\text{jic}}^+$  is monotonically increasing whereas  $\rho_{\text{jic}}^-$  is monotonically decreasing.  $\square$

### 2.7.3 Portfolio optimization models

For illustrating the use of various risk measures in practice, we present some portfolio optimization models. We consider a one-period financial portfolio optimization problem with  $n$  risky assets. Let  $\eta^T = (\eta_1, \dots, \eta_n)$  be the vector of random returns of the assets and  $r_i := \mathbb{E}[\eta_i]$ ,  $i = 1, \dots, n$  be the expected returns. The asset-weights in the portfolio will be denoted by  $x_1, \dots, x_n$ , thus  $\eta^T x$  represents the random portfolio return. Since for risk measures we have interpreted positive values of random variables as losses, we take  $\zeta(x, \eta) := -\eta^T x$ . With  $\mu_p$  standing in this section for a prescribed minimal expected portfolio return level, we consider optimization problems of the following form:

$$\left. \begin{aligned} \psi(\mu_p) &= \min \rho(-\eta^T x) \\ \text{s.t. } r^T x &\geq \mu_p \\ \mathbb{1}^T x &= 1 \\ x &\in \mathcal{B}, \end{aligned} \right\} \quad (2.206)$$

where  $\mathcal{B}$  is a polyhedral set determined by additional linear constraints,  $\rho$  is a risk measure, and  $\mathbb{1}^T = (1, \dots, 1)$  holds. The interpretation is the following: we are looking for a portfolio with minimum risk, under prescribing a minimum acceptable level  $\mu_p$  of expected portfolio return. This formulation of the portfolio selection problem is called a *risk–reward* model, with  $\rho(-\eta^T x)$  standing for risk and  $r^T x$  representing reward.

Some well-known particular cases, differing in the choice of the risk measure are the following:

- $\rho = \rho_{\text{Std}}$  corresponds to the classical minimum–variance model of Markowitz [217];
- $\rho = \rho_{\text{Std}}^-$  leads to the mean–semivariance model of Markowitz [217];
- $\rho = \rho_{\text{MAD}}$  gives the mean–absolute–deviation model of Konno and Yamazaki [194];
- $\rho = \rho_{\text{CVaR}}^\alpha$  corresponds to the mean–CVaR model of Rockafellar and Uryasev [282];
- $\rho = \rho_{\text{VaR}}^\alpha$  results in the mean–VaR model widely used in the finance industry, see, for instance, Jorion [151].

Note that all of these risk measures belong to the class of deviation measures. Although problem (2.206) is also useful in its own right, in finance this problem is considered as a parametric optimization problem with parameter  $\mu_p$ . The optimal objective value  $\psi(\mu_p)$ , as a function of  $\mu_p$ , plays an important role. Its graph in  $\mathbb{R}^2$  is called the *efficient frontier*, corresponding to the risk  $\rho$  and return  $\mu_p$ . Traditionally, the efficient frontier is represented graphically with the horizontal axis corresponding to risk and the vertical one corresponding to return.

The reason behind considering the efficiency curve is the following: we actually face a bi-objective optimization problem, where we would like to maximize the expected return and at the same time minimize risk. In all cases listed above,  $\psi(\mu_p)$  is strictly monotonically increasing in  $\mu_p$ , on the interval where the constraint  $r^T x \geq \mu_p$  is active at the optimal solution. Consequently, on the  $\psi$ –interval corresponding to this interval,  $\psi^{-1}$  exists and is strictly monotonically increasing. Thus it makes sense to consider the following alternative representation of the efficiency curve:

$$\left. \begin{aligned} \mu(\psi_p) &= \max r^T x \\ \text{s.t. } \rho(-\eta^T x) &\leq \psi_p \\ \mathbb{1}^T x &= 1 \\ x &\in \mathcal{B}, \end{aligned} \right\} \quad (2.207)$$



where now  $\psi_p > 0$  plays the role of a parameter. The interpretation of this problem is the following: we maximize expected return under the condition that the maximum acceptable risk is  $\psi_p$ . The portfolio optimization model in the above form is called a *reward–risk* model.

Due to the multi-objective character of the problem setting, it is not surprising that a third characterization of the efficient frontier is via the optimization problem

$$\left. \begin{array}{l} \max r^T x - v \rho(-\eta^T x) \\ \text{s.t. } \mathbf{1}^T x = 1 \\ x \in \mathcal{B}, \end{array} \right\} \quad (2.208)$$

where in this case  $v \geq 0$  is acting as a (risk–aversion) parameter for the efficiency curve. The evaluation function for the risk is accounted for by an additive term with a negative sign.

For details on the relationship between these three problems see, for instance, Palmquist, Uryasev, and Krokmal [249].

Let us finally remark that taking  $\zeta(x, \eta, \xi) := \eta^T x - \xi$  instead of  $\zeta(x, \eta) := \eta^T x$  also leads to an important class of portfolio optimization problems. In this case  $\xi$  may represent, for instance, the random return of a benchmark which can be, for instance, an index like the Dow Jones Industrial Average.

### 2.7.4 Optimizing performance

In this section we consider random variables of the form

$$\zeta(x, \eta, \xi) := \eta^T x - \xi$$

with positive values representing gains and negative values representing losses.

Gains will be measured via the expected value

$$f(x) := \mathbb{E}[\eta^T x - \xi] = \mu^T x - \mu_{n+1}$$

(termed in this context as *reward*), where we have employed the notation  $\mu := \mathbb{E}[\eta] \in \mathbb{R}^n$  and  $\mu_{n+1} := \mathbb{E}[\xi] \in \mathbb{R}$ . Losses will be measured via the evaluation function corresponding to a risk measure by

$$g(x) := \rho(-\eta^T x + \xi),$$

termed as *risk*. Notice that we have substituted  $\vartheta = -\zeta(x, \eta, \xi)$  into  $\rho(\cdot)$  in order to being in accordance with the convention in Section 2.7.1, where positive values of  $\vartheta$  represented losses.

As a performance measure we choose the reward–to–risk ratio

$$\frac{f(x)}{g(x)} = \frac{\mathbb{E}[\eta^T x - \xi]}{\rho(-\eta^T x + \xi)} = \frac{\mu^T x - \mu_{n+1}}{\rho(-\eta^T x + \xi)}$$

representing reward per unit risk.

Note that the selection of an appropriate performance measure heavily depends on the specific application. Considering, for example, financial portfolio optimization, there is a wide variety of performance evaluation measures in use, see e.g. Coigneau and Hubner [44]. In the financial portfolio optimization context (see Section 2.7.3), our performance measure corresponds to the classical Sharpe–ratio when choosing the standard deviation  $\rho_{\text{Std}}$  as the risk measure. The Sharpe–ratio is one of the most widely used performance measure in the field of financial portfolio management.

The performance optimization problem is formulated as

$$\left. \begin{array}{l} \max_x \frac{f(x)}{g(x)} \\ \text{s.t. } x \in \mathcal{B} \end{array} \right\} \equiv \left. \begin{array}{l} \max_x \frac{\mu^T x - \mu_{n+1}}{\rho(-\eta^T x + \xi)} \\ \text{s.t. } x \in \mathcal{B}, \end{array} \right\} \quad (2.209)$$

where  $\mathcal{B} = \{x \mid Ax = b, x \geq 0\}$  is a polyhedral set. Considering the financial portfolio optimization case, the solutions of (2.209) are called *tangential portfolios*. Concerning (2.209) we make the following assumptions:

- A1. The set  $\mathcal{B}$  of feasible solutions is nonempty and bounded.
- A2.  $g(x) > 0$  holds for all  $x \in \mathcal{B}$ .
- A3.  $\exists \hat{x} \in \mathcal{B}$  for which  $f(\hat{x}) > 0$  holds.
- A4.  $\rho$  is a positively homogeneous risk measure and  $g(x)$  is continuous.

Assumptions A1 and A4 imply that for (2.209) optimal solutions exist. Notice also that A2 and A3 immediately imply that at optimal solutions  $x^*$  the inequality  $\frac{f(x^*)}{g(x^*)} > 0$  must hold, that is,  $f(x^*) > 0$  holds.

From the optimization point of view problem (2.209) belongs to the class of fractional programming problems. Assuming that  $f(x) \geq 0$  holds for all  $x \in \mathcal{B}$  and that  $\rho$  is a convex risk measure, Proposition 2.32. on page 83 implies that the objective function of (2.209) is pseudo–concave, thus (2.209) has favorable properties from the numerical point of view.

It is a well–known fact in fractional programming that under appropriate assumptions (2.209) can be equivalently formulated as a convex programming problem, see e.g. Avriel et al. [8] and Schaible [298], [297]. Under the positive homogeneity assumption A4 concerning  $\rho$ , the equivalent convex programming problem can be further simplified, see Stoyanov, Rachev and Fabozzi [314].

For deriving the equivalent convex programming problem let us consider the following problem first:

$$\left. \begin{array}{ll} \max_{y,t} & \mu^T y - \mu_{n+1} t \\ \text{s.t.} & \rho(-\eta^T y + \xi t) = 1 \\ & Ay - bt = 0 \\ & y \geq 0 \\ & t \geq 0. \end{array} \right\} \quad (2.210)$$

We observe that for any  $(y, t)$  for which the last three constraints hold we have: if  $t = 0$  then  $y = 0$  follows. In fact, if  $t = 0$  and  $y \neq 0$  then  $Ay = 0$ ,  $y \geq 0$  has a nontrivial solution which contradicts the boundedness of  $\mathcal{B}$  (cf. assumption A1).

We will show that under our assumptions (2.209) and (2.210) are equivalent. In fact, let  $x$  be a feasible solution of (2.209). Then, due to the positive homogeneity of  $\rho$ , with

$$t := \frac{1}{g(x)} = \frac{1}{\rho(-\eta^T x + \xi)} > 0, \quad y := tx,$$

$(y, t)$  is a feasible solution of (2.210) and the corresponding objective function values of the two problems are equal.

Conversely, assume that  $(y, t)$  is a feasible solution of (2.210). Then  $t > 0$  must hold, because otherwise, as we have seen above,  $y = 0$  follows. But for  $(y, t) = (0, 0)$  the first constraint in (2.210) cannot hold since for the positively homogeneous  $\rho$  we have  $\rho(0) = 0$ . Thus, for any feasible solution  $t > 0$  holds. Then  $x := \frac{1}{t}y$  is obviously feasible in (2.209) and the positive homogeneity of  $\rho$  implies that the corresponding objective function values are equal.

Consequently, (2.209) and (2.210) are equivalent. In particular, (2.210) has an optimal solution  $(y^*, t^*)$  and the optimal objective values are equal. Since the optimal objective value of (2.209) is positive, we have

$$\mu^T y^* - \mu_{n+1} t^* > 0.$$

Next we consider the following relaxation of (2.210)

$$\left. \begin{array}{ll} \max_{y,t} & \mu^T y - \mu_{n+1} t \\ \text{s.t.} & \rho(-\eta^T y + \xi t) \leq 1 \\ & Ay - bt = 0 \\ & y \geq 0 \\ & t \geq 0, \end{array} \right\} \quad (2.211)$$

where the first equality constraint has been relaxed. We are going to prove that under our assumptions the equivalence of (2.210) and (2.211) follows.

Since (2.211) is a relaxation of (2.210), the optimal objective value of (2.211) must be positive provided that an optimal solution exists. Although  $y = 0$ ,

$t = 0$  is a feasible solution of (2.211), it cannot be optimal since the corresponding objective value is 0. In general, concerning optimality it is sufficient to consider feasible solutions of (2.211) with positive objective function values.

Let  $(y, t)$  be a feasible solution with  $\mu^T y - \mu_{n+1} t > 0$ . Assume that for such a solution the first inequality constraint in (2.211) is inactive, that is,  $\rho(-\eta^T y + \xi t) < 1$  holds. We take

$$\gamma := \frac{1}{\rho(-\eta^T y + \xi t)} > 1, \quad \bar{y} := \gamma y, \quad \bar{t} := \gamma t.$$

Then, due to the positive homogeneity of  $\rho$ ,  $(\bar{y}, \bar{t})$  is obviously a feasible solution of (2.211) with the first inequality constraint being active. For the corresponding objective function value we get:

$$\mu^T \bar{y} - \mu_{n+1} \bar{t} = \gamma(\mu^T y - \mu_{n+1} t) > \mu^T y - \mu_{n+1} t.$$

Consequently, with any feasible  $(y, t)$  for which  $\mu^T y - \mu_{n+1} t > 0$  holds and for which the first inequality constraint in (2.211) is inactive, we can associate a feasible solution  $(\bar{y}, \bar{t})$  having the following properties: it has a higher objective function value than  $(y, t)$  and for this feasible solution the first constraint is active in (2.211). This implies that problems (2.210) and (2.211) are equivalent, the optimal solution of (2.211) exists and the first constraint in (2.211) is active at the optimum.

We have shown that the following proposition holds:

**Proposition 2.53.** *Let us assume that A1–A4 hold. Then the fractional programming problem (2.209) is equivalent to (2.211).*

If  $\rho$  is a coherent risk measure, then (2.211) is clearly a convex programming problem, serving as an equivalent formulation of the original fractional programming problem (2.209).

For the equivalent reformulation the assumption  $g(x) > 0, \forall x \in \mathcal{B}$  is an essential one. From the modeling point of view, enforcing this could be done by adding the inequality  $g(x) \geq \varepsilon$  to the set of constraints of (2.209), with some  $\varepsilon > 0$ . The difficulty: if  $\rho$  is a convex risk measure, this is a reverse convex constraint, transforming (2.209) and its reformulation (2.211) into non-convex optimization problems. Therefore, we consider the following alternative formulation of (2.209):

$$\left. \begin{array}{ll} \min_x & \frac{g(x)}{f(x)} \\ \text{s.t.} & x \in \mathcal{B} \end{array} \right\} \equiv \left. \begin{array}{ll} \min_x & \frac{\rho(-\eta^T x + \xi)}{\mu^T x - \mu_{n+1}} \\ \text{s.t.} & x \in \mathcal{B}. \end{array} \right\} \quad (2.212)$$

If  $f(x) > 0$  and  $g(x) > 0$  hold for  $\forall x \in \mathcal{B}$  then this problem is clearly equivalent to (2.209). Therefore we keep assumptions A1, A2, A4 and replace A3 with the stronger requirement

A3'.  $f(x) > 0$  holds for all  $x \in \mathcal{B}$ .

If  $g$  is a convex function then (2.212) involves the minimization of a pseudo-convex objective function under linear constraints. For deriving the equivalent convex program we proceed analogously as for the case of (2.209), cf. Stoyanov, Rachev and Fabozzi [314]. We consider the problem

$$\left. \begin{array}{ll} \min_{y,t} & \rho(-\eta^T y + \xi t) \\ \text{s.t.} & \mu^T y - \mu_{n+1} t = 1 \\ & Ay - bt = 0 \\ & y \geq 0 \\ & t \geq 0. \end{array} \right\} \quad (2.213)$$

If  $x$  is a feasible solution of (2.212) then

$$t := \frac{1}{f(x)} = \frac{1}{\mu^T x - \mu_{n+1}} > 0, \quad y := tx,$$

is feasible for (2.213), due to the positive homogeneity of  $\rho$  and the corresponding objective values are equal. Conversely, let  $(y, t)$  be a feasible solution of (2.213) then  $t > 0$  must hold and  $x := \frac{1}{t}y$  is feasible for (2.212) with the same objective value. Thus, (2.212) and (2.213) are equivalent.

By relaxing the first equality constraint in (2.213) we get the following problem:

$$\left. \begin{array}{ll} \min_{y,t} & \rho(-\eta^T y + \xi t) \\ \text{s.t.} & \mu^T y - \mu_{n+1} t \geq 1 \\ & Ay - bt = 0 \\ & y \geq 0 \\ & t \geq 0. \end{array} \right\} \quad (2.214)$$

It is easy to see that the relaxed constraint must be active at any optimal solution of (2.214). In fact, let us assume that for a feasible solution  $(y, t)$  the inequality  $\mu^T y - \mu_{n+1} t > 1$  holds. Taking

$$\kappa := \frac{1}{\mu^T y - \mu_{n+1} t} < 1, \quad \bar{y} := \kappa y, \quad \bar{t} := \kappa t,$$

$(\bar{y}, \bar{t})$  is obviously feasible for (2.214) and the first inequality constraint becomes active. Utilizing the positive homogeneity of  $\rho$ , for the objective function values we get

$$\rho(-\eta^T \bar{y} + \xi \bar{t}) = \kappa \rho(-\eta^T y + \xi t) < \rho(-\eta^T y + \xi t),$$

implying that the first inequality constraint in (2.214) must be active at optimum. Consequently, the optimization problems (2.213) and (2.214) are equivalent.

The results above can be summarized in

**Proposition 2.54.** *Let us assume that A1, A2, A3' and A4 hold. Then the fractional programming problem (2.212) is equivalent to (2.214).*

If  $\rho$  is convex then (2.214) is clearly a convex programming problem.

Concerning (2.212), the requirement  $f(x) > 0$  for all  $x \in \mathcal{B}$  can be enforced by adding a linear constraint of the form  $\mu^T x - \mu_{n+1} \geq \varepsilon$  to the defining set of linear relations of  $\mathcal{B}$ , with a suitably chosen  $\varepsilon > 0$ .

As we have seen, under our assumptions the optimization problems (2.212), (2.213), (2.214) are equivalent formulations of the original fractional programming problem (2.209). Notice that for proving the equivalence we merely needed the positive homogeneity of  $\rho$ . Assuming additionally that  $\rho$  is subadditive, the equivalent problem (2.214) becomes a convex programming problem.

## Exercises

**2.12.** The portfolio optimization model of Young [352] corresponds to the choice of  $\rho_{\text{fat}}$  as a risk measure in our general portfolio optimization model (2.206). The idea is to select a portfolio which minimizes the maximum loss. We consider the case of a finite discrete distribution of the asset returns  $\eta$ :  $(p_i, \hat{\eta}^i)$ ,  $p_i = \mathbb{P}[\eta = \hat{\eta}^i]$ ,  $i = 1, \dots, N$ . Let  $r := \mathbb{E}[\eta]$ . The model formulation is:

$$\left. \begin{array}{l} \max_x \min_{1 \leq i \leq N} (\hat{\eta}^i)^T x \\ r^T x \geq \mu_p \\ \mathbb{1}^T x = 1 \\ x \geq 0 \end{array} \right\}$$

with  $\mathbb{1} = (1, \dots, 1)^T$ .

- (a) Explain why in the above model the maximum loss is minimized.
- (b) Give for this model an equivalent linear programming formulation.

**2.13.** A portfolio optimization problem is given as follows. There are two risky assets, with random returns  $\eta = (\eta_1, \eta_2)$ , where  $(\eta_1, \eta_2)$  has a finite discrete distribution with three realizations, given by the scenario tableau:

$p_k$	0.3	0.5	0.2
$\eta_1^k$	-0.003	0.02	0.01
$\eta_2^k$	0.06	-0.006	0.02

The minimum expected return is  $\mu_p = 0.018$ . As the risk measure to be minimized choose CVaR. With the random portfolio return  $\zeta(x, \eta) = \eta_1 x_1 + \eta_2 x_2$  this means that the objective function in (2.206) will be  $\rho_{\text{CVaR}}^\alpha(-\zeta(x, \eta))$  where we choose

$\alpha = 0.99$ . Set up the portfolio optimization problem and solve it by employing SLP-IOR.

**2.14.** Prove the following fact: if  $x^*$  is an optimal solution of (2.208) with  $v > 0$  then

- (a)  $x^*$  is an optimal solution of (2.206) with  $\mu_p = r^T x^*$  and
- (b) it is an optimal solution of (2.207) with  $\psi_p = \rho(-\eta^T x^*)$ .

Stochastic Linear Programming  
Models, Theory, and Computation

Kall, P.; Mayer, J.

2011, XX, 426 p., Hardcover

ISBN: 978-1-4419-7728-1