Technische Universität München
Zentrum Mathematik
HVB-Stiftungsinstitut für Finanzmathematik

Robust Optimization with Application in Asset Management

Katrin Schöttle

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:		Univ.-Prof. Dr. Claudia Klüppelberg
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Rudi Zagst
	2.	Dr. Jan-Joachim Rückmann, Senior Lecturer University of Birmingham / U.K.

Die Dissertation wurde am 27.06.2007 bei der Technischen Universität eingereicht und durch die Fakultät für Mathematik am 12.11.2007 angenommen.

Acknowledgements

First of all I would like to thank my advisor Prof. Dr. Rudi Zagst who offered me the chance to do a dissertation at the Institute for Mathematical Finance at the TU München. He provided a comfortable working environment and was open for questions whenever needed. I am also very grateful for the possibilities and the financial support that allowed me to present my research results at various international conferences. I would furthermore like to thank Dr. habil. Jan Rückmann for being my co-referee.

My most sincere thanks go to Dr. Ralf Werner, without whom this dissertation would not have been possible. Not only did he initiate the thesis topic, he also spent many evenings and long Sunday afternoons discussing mathematical problems with me. He furthermore answered patiently many questions and proposed valuable ideas. Thanks a lot for everything, Ralf!

Finally, I want to thank my colleagues, friends and family for patiently listening to my complaints about unfinished proofs or incomplete sections and for encouraging me throughout the time.

Abstract

This dissertation first reviews parametric convex conic optimization problems with respect to stability (continuity) properties. Afterwards, the general problem formulation gets modified using the robust counterpart approach of Ben-Tal and Nemirovski [4] to account for uncertainty in the parameters. This is done by introducing an uncertainty set for the parameters and performing a worst-case optimization. After analyzing the robust program as well with respect to stability, it is shown that robustification with an ellipsoidal uncertainty set leads to a unique and continuous optimal solution in many cases and the costs associated with such a benefit are qualified.

In the second part of the dissertation the robust counterpart approach is applied to the portfolio optimization problem of Markowitz [56] whose solution is known to be rather dependent on the input parameters. Hence, the main task in practice is to determine parameter estimates and to create appropriate uncertainty sets, especially around the vector of expected asset returns which crucially influences the outcome of the portfolio optimization problem. We illustrate different definitions of ellipsoidal uncertainty sets for the return vector and the consequences of the according robust optimization problems. Finally, consistency of parameters, uncertainty sets and of the resulting classical and robust portfolio estimates is investigated as well.

Zusammenfassung

In dieser Dissertation werden zunächst parametrische konvexe Kegeloptimierungsprobleme hinsichtlich ihrer Stabilitäts- bzw. Stetigkeitseigenschaften betrachtet. Anschließend wird die allgemeine Formulierung anhand des „robust counterpart“ Ansatzes von Ben-Tal und Nemirovski [4] modifiziert, um Parameterunsicherheit explizit berücksichtigen zu können. Dies wird dadurch erreicht, dass an Stelle eines konkreten Parameters eine Unsicherheitsmenge eingeführt wird und über den schlechtesten Fall optimiert wird. Nach der Analyse des robusten Optimierungsproblems hinsichtlich seiner Stetigkeitseigenschaften wird gezeigt, dass in vielen Fällen die Robustifizierung unter Verwendung einer elliptischen Unsicherheitsmenge zu einer eindeutigen und stetigen Lösung des Optimierungsproblems führt. Die auftretenden Kosten, die mit einem solchen Ansatz verbunden sind, werden ebenfalls untersucht und qualifiziert.

Im zweiten Teil der Dissertation wird dieser „robust counterpart“ Ansatz auf das Portfoliooptimierungsproblem von Markowitz [56] angewandt, von welchem bekannt ist, dass die Lösung sehr stark von den Inputparametern abhängt. Das Hauptproblem bei praktischen Fragestellungen ist demnach die Bestimmung von adäquaten Parametern und die Definition von geeigneten Unsicherheitsmengen, insbesondere für den Vektor der erwarteten Assetrenditen, welche das Resultat der Portfoliooptimierung maßgeblich beeinflussen. Es werden verschiedene elliptische Unsicherheitsmengen für den Renditevektor und die Konsequenzen der zugehörigen robusten Optimierungsprobleme dargestellt. Abschließend wird noch die Eigenschaft der Konsistenz für Parameterschätzer, Unsicherheitsmengen und die daraus resultierenden klassischen bzw. robusten Portfolios untersucht.

Contents

Chapter 1

Introduction

In most optimization programs in practical applications parameters are included that describe some objects that enter the general problem formulation. These parameters can e.g. represent values like length, volume, etc. in engineering applications, or they can describe characteristics of a financial market in asset management. In any case, those parameters affect the outcome of the optimization problem, and they need to be determined beforehand. They can either be measured (like lengths) or they have to be approximated or estimated from a sample containing similar data, as for example estimating the average asset returns from a historical data sample. Both methods can lead to inexact values implying that the optimization parameters are *uncertain*. The optimization problem can thus be solved by using a particular estimate for the parameter. It then is desirable that the optimal solution does not change too much if the parameter is modified a little, yielding stability of the optimization problem.

Besides using robust estimators as input parameters, there exist various approaches taking the uncertainty about the parameters explicitly into account when modelling the problem. This can for example be done by reformulating the optimization problem as a *stochastic program*. A different access to incorporating parameter uncertainty is given by the *robust counterpart approach* which was introduced in 1998 by Ben-Tal and Nemirovski [4]. Their idea basically is a worst-case approach as the optimal solution is determined such that it minimizes the worst possible outcome when all parameters lying within a so-called uncertainty set are considered. This robust counterpart approach is extensively studied in this dissertation, both from a theoretical point of view and when applied to a portfolio optimization problem from asset management.

As already mentioned, in the original parametric optimization, the unknown or uncertain parameter influences the solution of the optimization problem, and furthermore, some estimate is needed for being able to solve the problem in the first place. Similarly, in the robust counterpart approach, the uncertainty set describing possible parameter values crucially affects the solution of the optimization problem. And to reformulate and thus solve the robust formulation,

an explicit (and furthermore also manageable) definition of a practically relevant uncertainty set is needed. Hence, coming up with such an uncertainty set is not a trivial task.

1.1 Thesis organization

This dissertation is organized as follows. In Chapter 2 we introduce the general setting and the parametric convex conic optimization problem (GCP_u) that represents the foundation throughout all the investigations. After setting the notation we review existing continuity results for optimization problems and make use of them to obtain stability statements for our particular problem in the conic context. The main part there will be to determine conditions under which the set of solutions of the optimization problem is a singleton and furthermore continuous in the parameter u. Results about the feasibility set, the optimal value function and the set of ε-optimal solutions are given as well.

In Chapter 3 we present the robust counterpart approach and investigate the resulting robust problem as well with respect to stability. We will show that the modified objective function and the robust constraints are again continuous and convex. This enables us to apply the results from Chapter 2 and thus yields that robustification of a problem maintains the same stability characteristics as the original one.

The robust counterpart approach has gained a lot of interest since its first introduction, and by now there exists a variety of suggestions how to choose the needed uncertainty set. Ben-Tal and Nemirovski themselves propose to use an ellipsoid or an intersection of ellipsoids, many others favor interval uncertainty. Even though the approach is applied with different uncertainty sets in the literature, so far no one has investigated if and what influence the particular shape of the uncertainty set can have on the optimal solution and its stability, respectively. In Section 3.3 we study the two main shapes of interval and ellipsoidal uncertainty, and we find that under certain (rather general) conditions an ellipsoidal uncertainty set leads to a unique optimal solution which then is continuous with respect to the parameter u.

We furthermore examine the costs of robustification, measured in an increase of the optimal objective value compared to the value of the original problem. Ben-Tal and Nemirovski have proved in [4] for linear programs that the increase is linear in the size of the uncertainty set. We show the same result, but generalized to the convex conic setting.

After the theoretical investigations of the original and the robust program in Chapters 2 and 3, the rest of the dissertation is devoted to application thereof to portfolio optimization, a problem from finance. The aim is to find the optimal allocation for investing into a finite number of available assets. The parameters needed in the optimization problem are therefore the vector of expected returns

of the underlying assets and their covariance matrix.

In Chapter 4 we first summarize some distributional aspects which are needed to model the underlying market, and we also present different estimators for the uncertain parameters. We then introduce the classical portfolio optimization problem and illustrate why robustification is desirable in this application.

The robust version of the portfolio optimization problem is given thereafter in Chapter 5. As the robust problem can only be solved if the uncertainty set is explicitly specified, we discuss different possibilities for creating uncertainty sets in practice. We will find that the natural choice of using a confidence ellipsoid as an uncertainty set leads to the rather surprising result that the efficient frontier obtained from the robust portfolio optimization problem coincides with a part of the efficient frontier from the classical problem. This fact seems to be unknown so far.

To get more insight into the asymptotic characteristics of the parameter estimates, the uncertainty sets and the resulting optimal portfolios, we investigate in Chapter 6 all these figures with respect to consistency, the property of an estimate converging to the true value if the number of data used to obtain the estimator tends to infinity. We find that when using consistent parameter estimates in the optimization problem, both the resulting classical and robust optimal portfolios are consistent estimates for the portfolio that would be obtained when using the original – but unknown – market parameters.

Finally, Chapter 7 introduces concepts to determine parameter estimates and uncertainty sets based on the usual uncertainty of the parameters but additionally including external knowledge. The Bayes model (see e.g. Meucci [57]) is a quite well-known approach where a certain prior assumption is made (e.g. by some expert) and is then combined with information from a data sample to obtain the final estimate. A different model is given by the Black-Litterman approach [15] which gained interest in the finance community in the last years. There, rather arbitrary expert forecasts about the performance of the individual assets can be combined with an assumed market distribution.

1.2 Related literature

The risk evolving from using possibly incorrect parameter values in the optimization is often denoted by *estimation risk* in the literature as usually the value that is finally used is just an estimate (mostly based on a data sample) for the unknown true parameter.

There are various approaches to account for estimation risk especially in portfolio optimization. Early consideration of estimation risk can for example be found in Jobson and Korkie [44] and Best and Grauer [13]. For improvement of the results many papers propose to use parameter estimators other than the classical maximum likelihood as these seem to be more robust and thus reduce the

influence of estimation risk on the optimal portfolios. Jobson and Korkie [45] and also Efron and Morris [25] use Stein-type estimator, and others like e.g. Jorion [47] suggest Bayesian estimators which combine a traditional parameter estimate with external prior information. Robust estimators in general are considered in Perret-Gentil and Victoria-Feser [64].

A different approach towards *robustification* of the optimization problem and its solution was investigated by Lauprête, Samarov and Welsch [51]. Instead of the traditional procedure to separate the two steps of estimating the parameters and solving the optimization problem, they merged them and optimized the resulting portfolio estimate based on robust estimation routines for the parameters. A similar appraoch is taken also by Mori [62] using Stein-type estimators.

A further concept – rather well-known by now but not widely used due to potential patent conflicts – to incorporate estimation risk is *resampling* which was developed by Michaud [60] and is based on ideas of Jorion [48]. In that approach, distributional assumptions on the parameters are used to *resample* (i.e. draw random samples according to a given distribution) the optimization parameters, solve the optimization problem each time and finally average the respective optimal solutions. This yields a more averaged solution which – if resampled often enough – should not contain much estimation risk any more. The approach has also been criticized, see Scherer [75], but nevertheless it is rather easy to understand and can serve as a competitive comparison to other robustification results, see e.g. Schöttle and Werner [77].

Different types of optimization that also address the problem of estimation risk are stochastic programming and chance-constrained optimization which incorporate distributional assumptions and uncertainty about parameters into the optimization problem in terms of probabilistic constraint formulations. We will not pursue such an approach in this dissertation.

Finally, as already described above, there is the robust counterpart approach which was introduced by Ben-Tal and Nemirovski in 1998, see e.g. [4], and also independently by El-Ghaoui, Oustry and Lebret [26]. In this approach, an entire set of possible parameter realizations is used for the optimization, but no assumptions about the distribution of the unknown parameters is needed – as is the case for many other robustification approaches.

As it is well known in finance that the portfolio optimization problem strongly depends on the input parameters – especially the estimate for the vector of expected returns – the need for robustification is evident. Thus, by now, there are quite a few papers applying this robust counterpart in various ways to solve optimization problems in finance.

To name a few, Ben-Tal, Nemirovski and Margalit [6] use the robust approach to model and solve multi-period portfolio optimization with transaction costs. Goldfarb and Iyengar [33] apply the robust counterpart approach on a factor model for asset returns. The robustification technique is also exploited by Lutgens and Sturm [54] who extend it to optimizing portfolios including options. Tütüncü

and Koenig [79] consider the problem of finding robust asset allocations by solving the robust problem using a saddle-point algorithm, and Lobo [52] presents how the robust portfolio optimization problem using different uncertainty sets can be cast as a second order cone program (SOCP). Robust portfolio optimization is not only applicable in the mean-variance framework, but also when using different risk measures instead of the variance, e.g. the Value-at-Risk (VaR), as shown in El-Ghaoui, Oks and Oustry [27].

Recenctly published, the comprehensive book of Cornuejols and Tütüncü [22] contains both various aspects of optimization (including stochastic and robust optimization) and extensive applications in different fields of finance. It also includes many references relating to financial optimization. Further references can be found in the books of Meucci [57] and Scherer [76].

Part I

Theory of convex conic optimization and the robust counterpart approach

Chapter 2

The general convex conic optimization problem

2.1 Conic optimization

Before stating the generic convex conic optimization problem in the next section, we want to give a short introduction to conic optimization in general and present all necessary definitions, notational conventions and some useful statements. We will mostly follow the book of Boyd and Vandenberghe [19].

Definition 2.1.

(i) A set $K \subset \mathbb{R}^m$ is called a cone *if*

$$\forall x \in K, \forall \lambda \in \mathbb{R}, \lambda \geq 0 \quad \Rightarrow \quad \lambda x \in K.$$

(ii) A set $K \subset \mathbb{R}^m$ is a convex cone, if it is convex and a cone.

(iii) A cone $K \subset \mathbb{R}^m$ is called a proper cone *or* ordering cone *if it closed and convex, has non-empty interior and is pointed, meaning that*

$$x \in K, -x \in K \quad \Rightarrow \quad x = 0.$$

Definition 2.2. *Let $K \subset \mathbb{R}^m$ be an ordering cone. Then K defines a partial ordering on $\mathbb{R}^m$ by*

$$x \geq_K y \quad \Leftrightarrow \quad x - y \in K.$$

The cone K defining this relation is called the positive cone *in $\mathbb{R}^m$. Analogously we use the expression* negative cone *and the corresponding notation $x \leq_K y$.*
The associated strict partial odering is defined by

$$x >_K y \quad \Leftrightarrow \quad x - y \in \operatorname{int} K.$$

Note that according to the previous definition the following notations can be used equivalently:

$$\begin{aligned} x \leq_K 0 &\quad\Leftrightarrow\quad x \in -K \\ x <_K 0 &\quad\Leftrightarrow\quad x \in \operatorname{int}(-K) \\ x \geq_K 0 &\quad\Leftrightarrow\quad x \in K \\ x >_K 0 &\quad\Leftrightarrow\quad x \in \operatorname{int}(K). \end{aligned}$$

We will prefer the notation on the left side to express the close relation to the classical partial ordering on $\mathbb{R}^m$, as for a general understanding one can imagine K to represent $\mathbb{R}^m_+$, the non-negative orthant. In some cases though it might be more convenient to also use the notation on the right – for example in transitions from single points in K to entire ε-neighborhoods lying in K.

Some of the most common cones are presented in the following examples.

Example 2.3. *As already mentioned, the easiest cone is $K = \mathbb{R}_+$, the set of non-negative real numbers. There, the partial ordering "$\leq_K$" corresponds to the usual ordering "$\leq$".*

This interpretation can be generalized to arbitrary dimensions. When $K = \mathbb{R}^m_+$, i.e. the cone is described by the non-negative orthant, the associated partial ordering is the standard inequality "$\leq$" between vectors:

$$x, y \in \mathbb{R}^m, x \leq y \quad\Leftrightarrow\quad x_i \leq y_i \ \forall i = 1, \ldots, m.$$

Example 2.4. *The* Lorentz cone *or* second order cone *(sometimes also called "ice-cream cone") is defined by*

$$L^m := \left\{ x \in \mathbb{R}^m \mid x_m \geq \sqrt{\sum_{i=1}^{m-1} x_i^2} \right\}.$$

- *For $m = 1$, we have $L^1 = \mathbb{R}_+$.*
- *For $m = 2$, we get*

  $$L^2 := \left\{ x \in \mathbb{R}^2 \mid x_2 \geq |x_1| \right\}$$

  which is illustrated in Figure 2.1. It holds for example that the point $\begin{pmatrix} -1 \\ 1 \end{pmatrix} \in L^2$, or, using the equivalent notation, $\begin{pmatrix} -1 \\ 1 \end{pmatrix} \geq_{L^2} 0$.
- *For $m = 3$, we get*

  $$L^3 := \{ x \in \mathbb{R}^3 \mid x_3^2 \geq x_1^2 + x_2^2, x_3 \geq 0 \}.$$

  The three dimensional Lorentz cone L^3 is illustrated in Figure 2.2. which nicely motivates the name "ice-cream cone".

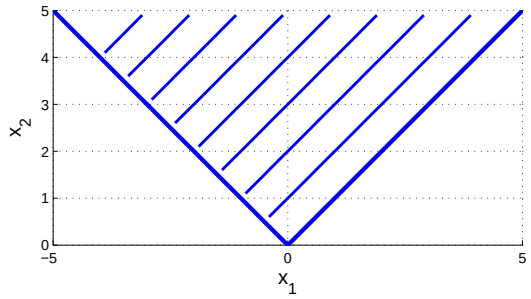


Figure 2.1: Illustration of the Lorentz cone L^2.

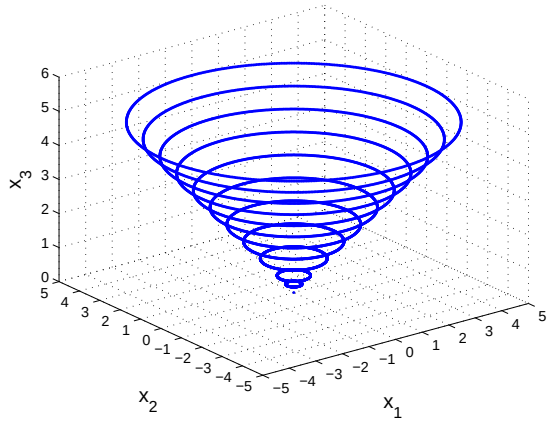


Figure 2.2: Illustration of the Lorentz cone L^3.

Example 2.5. *The space of symmetric $m \times m$ matrices is denoted by*

$$S^m = \{A \in \mathbb{R}^{m \times m} \mid A = A^T\}.$$

The cone of symmetric positive semidefinite matrices *describes a subset thereof and is defined by*

$$S^m_+ := \{A \in \mathbb{R}^{m \times m} \mid A = A^T,\ A \text{ positive semidefinite}\}.$$

The partial ordering "$\geq_{S^m_+}$" or simply "$\succeq$" ("$\leq_{S^m_+}$" or simply "$\preceq$") gives the characterization of matrices being positive (negative) semidefinite, i.e. for a symmetric

$m \times m$ matrix we have the following equivalent notations for $A \in S^m$:

$$\begin{aligned} A \in S^m_+ \quad &\Leftrightarrow \quad A \geq_{S^m_+} 0 \\ &\Leftrightarrow \quad A \succeq 0 \\ &\Leftrightarrow \quad A \text{ is positive semidefinite.} \end{aligned}$$

Analogous we define positive (negative) definiteness of a matrix A by $A \succ 0$ ($A \prec 0$). Similarly, the relation $A \preceq B$ ($A \prec B$) between two $m \times m$ matrices means that the difference matrix $B - A$ is positive semidefinite (positive definite).

We will also need the notion of the dual cone.

Definition 2.6. *The dual cone to the cone $K \subset \mathbb{R}^m$ is given by*

$$K^* := \{y \in \mathbb{R}^m \mid y^T x \geq 0 \ \ \forall x \in K\}.$$

Remark 2.7. *In a more general formulation, we use the inner product associated with the respective space to form the dual cone. In the definition above we were within the space $\mathbb{R}^m$, i.e. the standard inner product is given by $\langle x, y \rangle = x^T y$. In the space of symmetric matrices, S^m, the standard inner product is given by $\langle A, B \rangle = tr(A^T B) = tr(AB)$. Hence, the dual cone to the cone of symmetric positive definite matrices is defined as*

$$(S^m_+)^* = \{B \in S^m \mid tr(BA) \geq 0 \ \ \forall A \in S^m_+\}.$$

Note that this formula can also be derived by rewriting a matrix $A \in S^m$ as a vector $a \in \mathbb{R}^{m^2}$ (e.g. by stacking the columns underneath each other) and then applying the definition of the dual cone for vectors as given in Definition 2.6.

The dual cone K^* has some properties that are worth summarizing:

Proposition 2.8. *Let K be a cone.*

- *K^* is convex and closed.*
- *If K has non-empty interior, then K^* is pointed.*
- *If the closure of K is pointed, then K^* has non-empty interior.*
- *K^{**} is the closure of the convex hull of K. This especially implies $K^{**} = K$ if K is closed and convex, which holds e.g. for ordering cones.*

Proof. See e.g. Boyd and Vandenberghe [19], page 53. □

Remark 2.9. *The cones from Examples 2.3, 2.4 and 2.5, i.e. The cones $\mathbb{R}^m_+$, L^m and S^m_+ are* self-dual, *i.e. it holds $K = K^*$. We collect the respective proofs for completeness.*

- *For $K = \mathbb{R}^m_+$ the dual cone is given by*

$$\begin{aligned}(\mathbb{R}^m_+)^* &= \{y \in \mathbb{R}^m \mid y^T x \geq 0 \ \ \forall x \in \mathbb{R}^m_+\} \\ &= \{y \in \mathbb{R}^m \mid y_i \geq 0, i = 1, \ldots, m\} \\ &= \mathbb{R}^m_+.\end{aligned}$$

- *For $K = L^m$ the dual cone is given by*

$$(L^m)^* = \{y \in \mathbb{R}^m \mid y^T x \geq 0 \ \ \forall x \in L^m\}.$$

We prove the equality $(L^m)^* = L^m$ by showing both inclusions. Let the vector $(x^T, a)^T = (x_1, \ldots, x_{m-1}, a)^T$ be an arbitrary vector in the cone L^m, i.e. it holds that

$$a \geq \sqrt{\sum_{i=1}^{m-1} x_i^2} = \|x\|_2.$$

– *Let $(y^T, b)^T \in L^m$, i.e. $b \geq \|y\|_2$. Then we get, using the Cauchy-Schwarz inequality that*

$$\begin{aligned}(x^T, a) \begin{pmatrix} y \\ b \end{pmatrix} &= x^T y + ab \\ &\geq x^T y + \|x\|_2 \ \|y\|_2 \\ &\geq x^T y + |x^T y| \\ &\geq 0\end{aligned}$$

and thus, as $(x^T, a)^T \in L^m$ was arbitrary, $(y^T, b)^T \in (L^m)^*$.

– *Let $(y^T, b)^T \in (L^m)^*$, i.e. $x^T y + ab \geq 0$ for all $(x^T, a)^T \in L^m$. Thus, this especially holds for the vector $(x^T, a)^T = (-y^T, \|y\|_2)^T \in L^m$, i.e.*

$$(y^T, b) \begin{pmatrix} -y \\ \|y\|_2 \end{pmatrix} = -y^T y + \|y\|_2 b = \|y\|_2 \left(-\|y\|_2 + b\right) \geq 0$$

which implies, as $\|y\|_2 \geq 0$, that $b \geq \|y\|_2$, thus $(y^T, b)^T \in L^m$.

- *For $K = S^m_+$ the dual cone is given by (see Remark 2.7)*

$$(S^m_+)^* = \{B \in S^m \mid tr(BA) \geq 0 \ \ \forall A \in S^m_+\}.$$

We prove equality of $(S^m_+)^* = S^m_+$ again by showing both inclusions.

– *Let $B \in (S^m_+)^*$, i.e. B is symmetric and $tr(BA) \geq 0$ for all $A \in S^m_+$. We need to show positive semidefiniteness of B. Let $a \in \mathbb{R}^m$ arbitrary. Then the matrix $A := aa^T$ is positive semidefinite and thus in S^m_+.*

Using rules for matrix calculations, summarized in Appendix C, we obtain

$$a^T Ba = tr(a^T Ba) = tr(Baa^T) = tr(BA) \geq 0$$

according to the assumption. As $a \in R^m$ was arbitrary, this is an equivalent statement to B being positive semidefinite and thus $B \in S_+^m$.

– *Let $B \in S_+^m$, i.e. B is positive semidefinite and symmetric, and let $A \in S_+^m$ be arbitrary. Recall that any symmetric positive semidefinite matrix A can be decomposed as*

$$A = \sum_{i=1}^{m} \lambda_i v_i v_i^T$$

with $\lambda_i \geq 0$ being the eigenvalues and $v_i \in \mathbb{R}^m$ the eigenvectors of A, see e.g. [19], page 52. Thus, we get

$$\begin{aligned} tr(BA) &= tr\left(\sum_{i=1}^{m} \lambda_i B v_i v_i^T\right) \\ &= \sum_{i=1}^{m} \lambda_i tr\left(B v_i v_i^T\right) \\ &= \sum_{i=1}^{m} \lambda_i tr\left(v_i^T B v_i\right) \\ &= \sum_{i=1}^{m} \lambda_i \left(v_i^T B v_i\right) \\ &\geq 0, \end{aligned}$$

i.e. $B \in (S_+^m)^*$. The last inequality holds since B is positive semidefinite.

Subsequently, we will use the notion of the dual cone to give characterizations for elements lying within a cone $K \subset \mathbb{R}^m$ or within the interior $\operatorname{int} K$, respectively.

Lemma 2.10. *Let $K \subset \mathbb{R}^m$ be a closed convex cone. Then it holds that*

$$x \in K \quad \Leftrightarrow \quad \lambda^T x \geq 0 \quad \forall \lambda \in K^*.$$

Proof. The forward direction is obvious by the definition of the dual cone.

For the backward direction, we note that the dual cone of K^* is given by

$$K^{**} = \{z \in \mathbb{R}^m \mid \lambda^T z \geq 0 \ \ \forall \lambda \in K^*\}.$$

Thus, comparing K^{**} with the right hand side of the equivalence statement, we can conclude that $x \in K^{**}$. Furthermore, since K is closed, it holds that $K^{**} = K$ (Proposition 2.8), and thus $x \in K$. □

Lemma 2.11. *Let $K \subset \mathbb{R}^m$ be a closed convex cone with $\operatorname{int}(K) \neq \emptyset$. Then it holds that*

$$x \in \operatorname{int}(K) \quad \Leftrightarrow \quad \lambda^T x > 0 \quad \forall \lambda \in K^* \backslash \{0\}.$$

Proof. To prove the forward direction, let $x \in \operatorname{int}(K)$ and note that we already have

$$\lambda^T x \geq 0 \quad \forall \lambda \in K^*$$

from Lemma 2.10. It remains to show the strict inequality. Assume that there exists $\hat{\lambda} \in K^*, \hat{\lambda} \neq 0$ with $\hat{\lambda}^T x = 0$. Since $x \in \operatorname{int}(K)$, there exists an $\varepsilon > 0$ such that an entire ε-neighborhood of x is still contained within the interior of K, i.e.

$$V_\varepsilon(x) \subset \operatorname{int}(K).$$

Considering the point

$$y := x - \frac{\varepsilon}{2} \cdot \frac{\hat{\lambda}}{\|\hat{\lambda}\|_2} \in V_\varepsilon(x),$$

i.e. especially $y \in \operatorname{int}(K)$, we have

$$\hat{\lambda}^T y = \underbrace{\hat{\lambda}^T x}_{=0} - \frac{\varepsilon}{2} \cdot \underbrace{\frac{\hat{\lambda}^T \hat{\lambda}}{\|\hat{\lambda}\|_2}}_{=\|\hat{\lambda}\|_2 > 0} < 0$$

which is a contradiction to $\hat{\lambda}$ being within the dual cone K^*. Thus, it must hold that

$$\lambda^T x > 0 \quad \forall \lambda \in K^* \backslash \{0\}.$$

To prove the backward direction, we use the same argument as in Lemma 2.10 to obtain at least that $x \in K$. It remains to show that x lies in the interior. Assume that $x \in K, x \notin \operatorname{int}(K)$. Thus, as the complement of $\operatorname{int}(K)$ is closed, there exists a sequence $\{x_k\}, x_k \to x$ with $x_k \notin K$. Using Lemma 2.10 we know that there exist $\lambda_k \in K^* \backslash \{0\}$ with $\lambda_k^T x_k \leq 0$. Without loss of generality, let $\|\lambda_k\|_2 = 1$. These λ_k form a sequence on the compact set

$$S = \{\lambda \mid \lambda \in K^* \backslash \{0\}, \|\lambda\|_2 = 1\}.$$

Thus, there exist accumulation points within S, and without loss of generality (switch to a subsequence if necessary) we can assume that

$$\lambda_k \to \hat{\lambda} \in S.$$

Hence, in the limit we get

$$\lambda_k^T x_k \to \hat{\lambda}^T x \leq 0$$

which contradicts the prerequisite of $\lambda^T x > 0$ for all $\lambda \in K^* \backslash \{0\}$. Thus, we have

$$\lambda^T x_k > 0 \quad \forall \lambda \in K^* \backslash \{0\}.$$

Now we can use again the argument from Lemma 2.10 to conclude that all x_k must lie in $K^{**} = K$ which contradicts the assumption above. Hence, it must hold that $x \in \text{int}(K)$. □

As we want to investigate convex conic optimization problems in a general setting, we need to extend the notion of convexity of real-valued functions to convexity in the conic sense.

Definition 2.12. *Let $K \subset \mathbb{R}^m$ be a positive cone. A mapping $g : \mathbb{R}^l \to \mathbb{R}^m$ is said to be* K-convex *if*

$$g(\alpha x + (1-\alpha)y) \leq_K \alpha g(x) + (1-\alpha)g(y)$$

for all $x, y \in \mathbb{R}^l$ and all $\alpha, 0 \leq \alpha \leq 1$.
Analogously, strict K-convexity *of g is given if*

$$g(\alpha x + (1-\alpha)y) <_K \alpha g(x) + (1-\alpha)g(y)$$

for all $x, y \in \mathbb{R}^l$ and all $\alpha, 0 < \alpha < 1$.

Using the definition of the dual cone we can state the following result.

Proposition 2.13. *Let $K \subset \mathbb{R}^m$ be a positive cone. A mapping $g : \mathbb{R}^l \to \mathbb{R}^m$ is K-convex if and only if the real-valued function $\lambda^T g : \mathbb{R}^l \to \mathbb{R}$ is convex for all $\lambda \in K^*$, i.e. for all elements of the dual cone of K.*

Proof. For a proof see Bonnans and Shapiro [18], Section 2.3.5. □

2.2 General convex conic optimization problem

This section describes the type of optimization problem that will be considered throughout the rest of this dissertation. We will also use this introductary section to explain the notation and give general definitions that will be used.

If not stated otherwise, we will always assume the following:

Assumption 2.14.

- *The set $X \subset \mathbb{R}^n$ is non-empty, convex and compact.*
- *The parameter $u \in \mathcal{U}$ represents the vector of uncertain data.*
- *The set $\mathcal{U} \subset \mathbb{R}^d$ is non-empty, convex and compact.*

- $K \subset \mathbb{R}^m$ *is an ordering cone with* $\operatorname{int}(K) \neq \emptyset$.
- $f : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ *is continuous both in* x *and* u*, and* f *is convex in* u *for fixed* $x \in X$ *and convex in* x *for fixed* $u \in \mathcal{U}$.
- $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^m$ *is continuous both in* x *and* u*, and* g *is* K*-convex in* u *for fixed* $x \in X$ *and* K*-convex in* x *for fixed* $u \in \mathcal{U}$.

In this context $\mathcal{U}$ is called *uncertainty set* and contains all possible realizations of the (uncertain) parameter vector u. The assumption of $\mathcal{U}$ being compact is not too restrictive since in practical problems uncertainty sets are mostly defined in cases where a parameter cannot be measured or estimated exactly, but a rough value is usually known. Thus, boundedness is assured, and closedness can simply be assumed without loss of generality as uncertainty sets for a parameter u are usually chosen to be some sphere (e.g. ellipse or polytope or an intersection thereof) centered around a particular estimate of the unknown parameter.

Assuming (K-)convexity of f and g in the parameter u seems to be rather natural, since if a point x is feasible for two different parameters u_1 and u_2, it should also be feasible for the parameters in between. This does not have to be the case for concave functions. Furthermore, in Chapter 3 when considering a robust optimization problem, we will have to find a worst case parameter within the uncertainty set, i.e. maximizing some function over the set. It is more intuitive if the worst case parameter is attained at the boundary – which is the case when maximizing a convex function – and not in the middle.

The set X represents the set of constraints that do not depend on the uncertain parameter u. In practice it is usually described by a few inequalities (and equalities), e.g. in portfolio theory we will often have

$$X = \{x \in \mathbb{R}^n \mid Ax \leq b\}.$$

And as unbounded solutions of an optimization problem are not really reasonable in practice, requiring compactness for the set X does not really constitute a restriction.

After having specified the assumptions, we can now define the general optimization problem that will be the foundation for all investigations and applications in this dissertation.

Definition 2.15. *Let Assumption 2.14 hold. The general convex optimization problem depending on a given parameter* $u \in \mathcal{U}$ *will be referred to by* (GCP_u) *and is assumed to be stated as:*

$$\begin{aligned} \min_{x \in X} \quad & f(x,u) \\ s.t. \quad & g(x,u) \leq_K 0. \end{aligned} \tag{GCP_u}$$

As the general problem formulation depends on the particularly chosen parameter $u \in \mathcal{U}$, the feasibility set, the optimal value function and the set of optimal solutions will naturally be given in terms of u as well.

Definition 2.16.

(i) The feasible set mapping $\mathcal{F} : \mathcal{U} \to \mathcal{P}(X)$ *(with $\mathcal{P}(X)$ denoting the power set of X, see e.g. Jänich [43]) is the mapping assigning to each parameter $u \in \mathcal{U}$ the corresponding feasibility set of the problem (GCP_u). For given $u \in \mathcal{U}$ the feasibility set of (GCP_u) will be denoted by*

$$\mathcal{F}(u) := \{x \in X \mid g(x,u) \leq_K 0\}.$$

The set of points that are feasible for all $u \in \mathcal{U}$ will be denoted by

$$\mathcal{F}_{\mathcal{U}} := \bigcap_{u \in \mathcal{U}} \mathcal{F}(u).$$

(ii) The set of Slater points *of (GCP_u) will be denoted by $\mathcal{F}^S(u)$ and is given by*

$$\mathcal{F}^S(u) := \{x \in X \mid g(x,u) <_K 0\}$$

(iii) The extreme value function *or* optimal value function *will be denoted by $f^* : \mathcal{U} \to \mathbb{R}$ and is defined as*

$$f^*(u) := \min\{f(x,u) \mid x \in \mathcal{F}(u)\}.$$

(iv) The optimal set mapping $\mathcal{F}^* : \mathcal{U} \to \mathcal{P}(X)$ *is the mapping assigning to each parameter $u \in \mathcal{U}$ the set of all x that are optimal solutions to the program (GCP_u). The set of optimal solutions is thus defined as*

$$\mathcal{F}^*(u) := \{x \in \mathcal{F}(u) \mid f(x,u) \leq f^*(u)\}.$$

(v) The ε-optimal set mapping *is defined analogously through the set of ε-optimal solutions for the respective parameter $u \in \mathcal{U}$:*

$$\mathcal{F}^*_\varepsilon(u) := \{x \in \mathcal{F}(u) \mid f(x,u) \leq f^*(u) + \varepsilon\}.$$

Note that such mappings as e.g. $\mathcal{F}$ and $\mathcal{F}^*$ assigning a (possibly empty) subset of X to each element $u \in \mathcal{U}$ will in the following be referred to as point-to-set mappings, set-valued mappings or multi-valued mappings.

Subsequently, we furthermore make the general assumption that there exists at least one point which is feasible for all possible parameter realizations within the uncertainty set. This requirement is rather naturally fulfilled in most practical problems.

Assumption 2.17. $\mathcal{F}_{\mathcal{U}} \neq \emptyset$.

Remark 2.18. *Note that by making the assumption that $\mathcal{F}_{\mathcal{U}} \neq \emptyset$, we especially have a non-empty feasibility set of problem (GCP_u) for any parameter $u \in \mathcal{U}$. Furthermore, the feasibility set $\mathcal{F}(u)$ is compact as an intersection of a compact and a closed set. Non-emptiness and compactness of $\mathcal{F}(u)$ together with continuity of f in x hence assure the existence of an optimal solution of (GCP_u).*

In Definition 2.16 we have defined a whole set of optimal solutions for a given parameter $u \in \mathcal{U}$. This set $\mathcal{F}^*(u)$ could in general be empty (which will not be the case in our setting, as we minimize a continuous function over a compact set), contain exactly one optimal solution $x^*(u)$ or consist of finitely or infinitely many optimal solutions. Having an entire set of solutions, it can be of interest to select exactly one solution for each parameter. To mathematically deal with the notion of a selection, we make the following definition.

Definition 2.19.

(i) *A* selection $\gamma : \mathcal{U} \to X$ *of a multi-valued mapping*[1] Γ *is a mapping from the set of parameter realizations $\mathcal{U}$ onto* single elements *of the image set $\Gamma(u)$ i.e. $\gamma(u) \in \Gamma(u)$.*

(ii) *A selection $\zeta^* : \mathcal{U} \to X$ of the optimal set mapping $\mathcal{F}^*$ is therefore a mapping assigning a single optimal solution $\zeta^*(u) = x^*(u) \in \mathcal{F}^*(u)$ to each parameter $u \in \mathcal{U}$. A selection function within $\mathcal{F}^*_\varepsilon$ is defined analogously by ζ^*_ε.*

After having presented the general problem setting and notation, we are now interested in *stability* or *robustness* of the optimization problem and especially its optimal solution(s). With robustness – or equivalently stability – of a solution with respect to the (uncertain) parameter we basically mean that the optimal solution of an optimization problem should not change very much if the parameter u is only disturbed a little.

2.3 $\mathcal{U}$-stability

In this section we want to present and discuss a notion of stability of an optimization problem that reflects the desired properties for practical applications. The main goal of practitioners can be described with the following statements which also characterize a *well-posed* problem in the sense of Hadamard (nicely summarized in Kirsch [49], originally introduced in Hadamard [36]):

[1]The mapping $\Gamma : \mathcal{U} \to \mathcal{P}(X)$ describes a general multi-valued mapping from the set of parameters to a subset of X. This can e.g. represent the feasible set mapping $\mathcal{F}$ or the optimal set mapping $\mathcal{F}^*$.

(i) **Existence**
There exists at least one solution of the problem.

(ii) **Uniqueness**
For each parameter choice $u \in \mathcal{U}$ there exists exactly one solution of the considered problem.

(iii) **Stability**
When the parameter u is disturbed only slightly, the optimal solution should also change only very little, i.e. the solution depends continuously on the data.

In this section we want to discuss these statements for the generic convex program (GCP_u), and we will find that especially the last one, the stability of the solution with respect to the parameter, only holds under certain regularity conditions.

Beforehand, note that investigating uniqueness of a solution is only meaningful if the existence is already established. Similarly, if uniqueness is not yet assured, there is no point in examining stability of *the* optimal solution. We can then only discuss stability or continuity of the set of optimal solutions as a whole.

In our setting existence of a solution is assured since the objective function f was assumed to be continuous and the feasibility set $\mathcal{F}(u)$ is non-empty (Assumption 2.17) and compact for all possible parameter choices $u \in \mathcal{U}$. Thus, $\mathcal{F}^*(u) \neq \emptyset$ for all $u \in \mathcal{U}$.

Uniqueness of a solution means that the optimal set is a singleton, i.e. $\mathcal{F}^*(u) = \{x^*(u)\}$. In this case we can then use the classical notion of continuity of a function $x^* : \mathcal{U} \subset \mathbb{R}^d \to \mathbb{R}^n$. From optimization theory it is known that uniqueness of the solution of our optimization problem (GCP_u) can be guaranteed if the objective function $f(\cdot, u)$ is strictly convex for each $u \in \mathcal{U}$. Furthermore, we will find in this section that in this case we will get stability of the solution – that is, continuity with respect to the parameter – by requiring the existence of a Slater point. Thus, these two additional facts would result in a so-called well-posed problem fulfilling all the properties practitioners desire.

But, in general, we cannot assume that the optimization problem has a unique solution. Therefore, we need to deal with entire sets, and thus a different notion of continuity of sets is necessary which will be given by the Hausdorff continuity.

Recalling the three characteristics stated above, we thus want to find a single-valued continuous mapping from the set of (uncertain) parameters into the set of optimal solutions. We will call a program having this nice property to be $\mathcal{U}$*-stable*:

Definition 2.20. *The problem $(GCP_{\hat{u}})$ is called* $\mathcal{U}$-stable *if*

(i) the set of optimal solutions contains exactly one element, $\mathcal{F}^*(\hat{u}) = \{x^*(\hat{u})\}$,

(ii) the mapping x^* is continuous at $\hat{u}$.

Extending this definition of stability of a particular problem to the entire family of parametric programs leads to the following:

Definition 2.21. *The family of problems (GCP_u) is called* $\mathcal{U}$-stable *if the individual problem $(GCP_{\hat{u}})$ is $\mathcal{U}$-stable for each $\hat{u} \in \mathcal{U}$, i.e. if it holds that*

(i) the set of optimal solutions for each parameter contains exactly one element, i.e. $\mathcal{F}^*(u) = \{x^*(u)\}$ for all $u \in \mathcal{U}$,

(ii) the mapping x^* is continuous on $\mathcal{U}$.

Remark 2.22. *A $\mathcal{U}$-stable problem will sometimes equivalently be referred to as a well-posed problem.*

In the following subsections we will investigate the question under which (additional) requirements the optimization problem (GCP_u) will be $\mathcal{U}$-stable.

2.3.1 Review of existing results

Before starting with theoretical examinations, we need some more definitions for multi-valued mappings which in our context will be the mappings from the parameter set into the sets of feasible and (ε-) optimal solutions, i.e. the functions $\mathcal{F}, \mathcal{F}^*$ and $\mathcal{F}^*_\varepsilon$. The following definitions are according to the book of Bank et al. [3] as well as several results.

An *ε-neighborhood* ($\varepsilon > 0$) of a set $S \subset \mathbb{R}^n$ will be described by the expression

$$V_\varepsilon(S) := \{x \in \mathbb{R}^n \mid d(x, S) = \inf_{y \in S} d(x, y) < \varepsilon\}.$$

The distinction to an ε-neighborhood around a point $u_0 \in \mathbb{R}^d$ is made clear in the context, as the notation is analogously given by $V_\varepsilon(u_0)$.

Definition 2.23. *A point-to-set mapping $\Gamma : \mathcal{U} \to \mathcal{P}(\mathbb{R}^n)$ is*

(i) closed *at a point $\hat{u}$ if for each pair of sequences $\{u_k\} \subset \mathcal{U}$ and $\{x_k\} \subset \mathbb{R}^n, k = 1, 2, \ldots,$ with the properties $u_k \to \hat{u}, \quad x_k \in \Gamma(u_k), \quad x_k \to \hat{x}$, it follows that $\hat{x} \in \Gamma(\hat{u})$;*

(ii) Hausdorff upper semicontinuous *(short: H-usc) at a point $\hat{u}$ if for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $\Gamma(u) \subset V_\varepsilon(\Gamma(\hat{u}))$ for all $u \in V_\delta(\hat{u})$;*

(iii) Hausdorff lower semicontinuous *(short: H-lsc) at a point $\hat{u}$ if for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $\Gamma(\hat{u}) \subset V_\varepsilon(\Gamma(u))$ for all $u \in V_\delta(\hat{u})$;*

(iv) H-continuous *at $\hat{u}$ if it is H-usc and H-lsc at $\hat{u}$;*

(v) strongly lower semicontinuous *(short: strongly lsc) at a point $\hat{u}$ if for each $x \in \Gamma(\hat{u})$ there exists an $\varepsilon > 0$ and a $\delta > 0$ such that $V_\varepsilon(x) \subset \Gamma(u)$ for all $u \in V_\delta(\hat{u})$.*

This notion of Hausdorff continuity for sets matches the "classical" continuity for vectors in $\mathbb{R}^n$. As we will use several results stated in the book of Bank et al. [3], and many of them are expressed using a second type of continuity for set-valued mappings, we need to give the according definitions as well.

Definition 2.24. *A point-to-set mapping $\Gamma : \mathcal{U} \to \mathcal{P}(\mathbb{R}^n)$ is*

(i) Berge upper semicontinuous *(short: B-usc) at a point $\hat{u}$ if for each open set Ω containing $\Gamma(\hat{u})$ there exists a $\delta = \delta(\Omega) > 0$ such that $\Gamma(u) \subset \Omega$ for all $u \in V_\delta(\hat{u})$;*

(ii) Berge lower semicontinuous *(short: B-lsc) at a point $\hat{u}$ if for each open set Ω satisfying $\Omega \cap \Gamma(\hat{u}) \neq \emptyset$ there exists a $\delta = \delta(\Omega) > 0$ such that $\Gamma(u) \cap \Omega \neq \emptyset$ for all $u \in V_\delta(\hat{u})$;*

(iii) B-continuous *at $\hat{u}$ if it is B-usc and B-lsc at $\hat{u}$;*

Remark 2.25. *The following implications hold, see Bank et al. [3], page 26:*

- *B-usc $\implies$ H-usc,*
- *H-lsc $\implies$ B-lsc,*
- *strongly lsc $\implies$ B-lsc.*

A rather useful result stating that the backward directions H-usc $\implies$ B-usc and B-lsc $\implies$ H-lsc also hold if additionally some compactness conditions are fulfilled is the following:

Lemma 2.26. *Let $\Gamma : \mathcal{U} \to \mathcal{P}(X)$ be a multi-valued mapping and let $\hat{u} \in \mathcal{U}$.*

(i) Γ is B-usc and thus as well H-usc at $\hat{u}$ if Γ is closed at $\hat{u}$ and X is compact.

(ii) Γ is B-usc at $\hat{u}$ if Γ is H-usc at $\hat{u}$ and $\Gamma(\hat{u})$ is compact.

(iii) Γ is H-lsc at $\hat{u}$ if Γ is B-lsc at $\hat{u}$ and $\operatorname{cl}\Gamma(\hat{u})$ is compact.

Proof. See Bank et al. [3], Lemma 2.2.3. □

Thus, Remark 2.25 together with Lemma 2.26 give equivalence of Hausdorff and Berge continuity in case of dealing with compact sets. As we are interested in stability investigations of the mappings $\mathcal{F}$, $\mathcal{F}^*$ and $\mathcal{F}^*_\varepsilon$ where the corresponding sets are all compact, we can hence use the two continuity definitions equivalently.

Foreclosing some results, we will find that in case of the optimal solution being unique, it is also continuous with respect to the parameter u. But in the more general case of having an entire set of optimal solutions, not much can be proved about stability. Therefore, we are as well interested in the possibility of choosing or selecting one particular optimal point $x^*(u)$ for each parameter u with the property that this selection then is continuous. The link to such a continuous selection function is established by Berge lower semicontinuity:

- Bank et al. [3] prove in their Corollary 2.3.2.1 that in our setting there exists a continuous selection function within a set-valued mapping if this mapping is Berge lower semicontinuous.
  This statement is basically Michael's *selection theorem*. More details and results about continuous selections under various settings can e.g. be found in Repovš and Semenov [68].

- Theorem 0.44 in Repovš and Semenov [68] basically gives the other direction of the selection theorem, i.e. if there exists a locally continuous selection function, then the multi-valued mapping is Berge lower semicontinuous.

Thus, it suffices to prove Hausdorff lower semicontinuity of (ε-)optimal set mapppings to assure at least the existence of a continuous selection function within the general mapping.

For notational ease in some of the subsequent results or proofs, we make the following definition.

Definition 2.27.

(i) The set of points satisfying the parametric constraints will be denoted with $\mathcal{G}(u)$ and is defined as

$$\mathcal{G}(u) := \{x \in \mathbb{R}^n \mid g(x,u) \in -K\} = \{x \in \mathbb{R}^n \mid g(x,u) \leq_K 0\}.$$

(ii) The set of Slater points of $\mathcal{G}(u)$ for a parameter $u \in \mathcal{U}$ will be denoted by

$$\mathcal{G}^S(u) := \{x \in \mathbb{R}^n \mid g(x,u) \in \operatorname{int}(-K)\} = \{x \in \mathbb{R}^n \mid g(x,u) <_K 0\}.$$

Note that the only difference between $\mathcal{G}(u)$ and $\mathcal{F}(u)$ lies in the expression $x \in \mathbb{R}^n$ compared to $x \in X$, thus, we can express the feasibility set and the set of Slater points as

$$\begin{aligned} \mathcal{F}(u) &= \mathcal{G}(u) \cap X, \\ \mathcal{F}^S(u) &= \mathcal{G}^S(u) \cap X. \end{aligned}$$

Furthermore, it obviously holds that

$$\mathcal{G}^S(u) \subset \mathcal{G}(u).$$

A useful result linking the two sets $\mathcal{G}^S(u)$ and $\mathcal{G}(u)$ is given in the next proposition.

Proposition 2.28. *Let $\mathcal{G}(u)$ and $\mathcal{G}^S(u)$ be as defined above and assume further that $\mathcal{G}^S(u) \neq \emptyset$. Then it holds that*

$$\mathcal{G}(u) = \operatorname{cl} \mathcal{G}^S(u)$$

with $\operatorname{cl} A$ denoting the closure of the set A.

Proof. We will show equivalence of the two sets by showing that each side contains the other.

- The one direction is straightforward: We have that $\mathcal{G}^S(u) \subset \mathcal{G}(u)$, and since $\mathcal{G}(u)$ is closed, we get that

$$\operatorname{cl} \mathcal{G}^S(u) \subset \operatorname{cl} \mathcal{G}(u) = \mathcal{G}(u).$$

- Since by assumption $\mathcal{G}^S(u)$ is non-empty, there exists a point $x \in \mathcal{G}^S(u)$, i.e. $g(x,u) <_K 0$. Let $y \in \mathcal{G}(u)$ and define the point $x_\lambda := \lambda x + (1-\lambda)y$ with $\lambda \in (0,1)$. Since $g(\cdot, u)$ is K-convex, it holds that

$$g(x_\lambda, u) = g(\lambda x + (1-\lambda)y, u) \leq_K \lambda \underbrace{g(x,u)}_{<_K 0} + (1-\lambda) \underbrace{g(y,u)}_{\leq_K 0} <_K 0.$$

  Therefore, $x_\lambda \in \mathcal{G}^S(u)$ for all $\lambda \in (0,1)$ and thus in the limit $\lambda \to 0$ we get $y = x_0 \in \operatorname{cl} \mathcal{G}^S(u)$, hence $\mathcal{G}(u) \subset \operatorname{cl} \mathcal{G}^S(u)$, which completes the proof. □

Having all these definitions we want to study the characteristics of the different sets and mappings we have defined in Section 2.2, always keeping in mind that the main goal would be to define conditions such that the set of optimal solutions is single-valued and continuous or that there exists at least a continuous selection within the set. Since this might not be possible, we consider as well the set of ε-optimal solutions with respect to finding some continuity results.

We first summarize and extend selected existing results from Bank et al. [3] which we will use afterwards for proving the desired statements in our general convex conic setting.

Theorem 2.29. *Let the mapping $\mathcal{F}$ be closed at $\hat{u}$. Then $\mathcal{F}^*$ is closed at $\hat{u}$ if f^* is upper semicontinuous at $\hat{u}$ and f is lower semicontinuous on $X \times \{\hat{u}\}$.*

Proof. See Bank et al. [3], Theorem 4.2.1 (3). □

Corollary 2.30. *Let $\mathcal{F}$ be closed at $\hat{u}$, $\mathcal{F}(\hat{u})$ be non-empty, f be continuous, and X be compact. Then f^* is lower semicontinuous at $\hat{u}$; f^* is also upper semicontinuous at $\hat{u}$ if and only if $\mathcal{F}^*$ is B-usc at $\hat{u}$.*

Proof. See Bank et al. [3], Corollary 4.2.1.1. □

Theorem 2.31.

(i) The optimal value function f^* is upper semicontinuous at $\hat{u}$ if $\mathcal{F}$ is B-lsc at $\hat{u}$ and f is upper semicontinuous on $\mathcal{F}(\hat{u}) \times \{\hat{u}\}$.

(ii) The optimal value function f^* is lower semicontinuous at $\hat{u}$ if $\mathcal{F}$ is H-usc at $\hat{u}$, $\mathcal{F}(\hat{u})$ is compact and f is lower semicontinuous on $\mathcal{F}(\hat{u}) \times \{\hat{u}\}$.

Proof. See Bank et al. [3], Theorem 4.2.2 (1) and (2). □

Corollary 2.32. *If the mapping $\mathcal{F}$ is B-lsc at $\hat{u}$, then the following statements are equivalent:*

(i) f^* is continuous at $\hat{u}$;

(ii) The mapping $\widetilde{\mathcal{F}}_\varepsilon^*$ defined through

$$\widetilde{\mathcal{F}}_\varepsilon^*(u) := \{x \in \mathcal{F}(u) \mid f(x,u) < f^*(u) + \varepsilon\}$$

is B-lsc and H-lsc at $\hat{u}$ for each $\varepsilon > 0$.

Proof. See Bank et al. [3], Corollary 4.2.4.1. Hausdorff lower semicontinuity of $\widetilde{\mathcal{F}}_\varepsilon^*$ finally follows from Lemma 2.26, part (iii). □

The following lemma is a more technical result which is needed to prove the statement of Theorem 2.34 below.

Lemma 2.33. *Let $\Gamma_1, \Gamma_2, \Gamma_3$ be mappings from $\mathcal{U}$ into $\mathcal{P}(\mathbb{R}^n)$ with the properties*

(i) Γ_1 is B-lsc at $\hat{u}$,

(ii) Γ_2 is strongly lsc at $\hat{u}$,

(iii) $\Gamma_2(u) \subset \Gamma_3(u) \quad \forall u \in \mathcal{U}$.

Then $\Gamma_1 \cap \Gamma_3$ is B-lsc at $\hat{u}$ if $\Gamma_1(\hat{u}) \cap \Gamma_3(\hat{u}) \subset \mathrm{cl}(\Gamma_1(\hat{u}) \cap \Gamma_2(\hat{u}))$ holds.

Proof. See Bank et al. [3], Corollary 2.2.5.1. □

Theorem 2.34. *Let the mappings $\mathcal{G}$ and $\mathcal{G}^S$ be as denoted above and let $\Gamma : \mathcal{U} \to \mathcal{P}(\mathbb{R}^n)$ be B-lsc at $\hat{u}$. Additionally, let*

$$(\mathcal{G} \cap \Gamma)(\hat{u}) \subset \mathrm{cl}(\mathcal{G}^S \cap \Gamma)(\hat{u}).$$

Then the mapping $\mathcal{G} \cap \Gamma$ is B-lsc at $\hat{u}$.

Proof. To prove the theorem, we apply Lemma 2.33 with appropriate assignments of the mappings Γ_1, Γ_2 and Γ_3.

- We choose $\Gamma_1 = \Gamma$ which then is B-lsc by assumption.
- Defining $\Gamma_2 = \mathcal{G}^S$, it remains to show that $\mathcal{G}^S$ is strongly lsc.
- With $\Gamma_3 = \mathcal{G}$, we obviously have $\Gamma_2(u) \subset \Gamma_3(u) \quad \forall u \in \mathcal{U}$. And we also have – by assumption in the theorem – that $\Gamma_1(\hat{u}) \cap \Gamma_3(\hat{u}) \subset \mathrm{cl}(\Gamma_1(\hat{u}) \cap \Gamma_2(\hat{u}))$.

Therefore, for applying Lemma 2.33 and thus proving that $\Gamma_1 \cap \Gamma_3 = \Gamma \cap \mathcal{G}$ is B-lsc at $\hat{u}$, it suffices to show strong lower semicontinuity of $\mathcal{G}^S$:
Let $\hat{x} \in \mathcal{G}^S(\hat{u})$, i.e. $g(\hat{x}, \hat{u}) <_K 0$. Since g is continuous on $X \times \mathcal{U}$, there exist $\varepsilon > 0$ and $\delta > 0$ such that $V_\varepsilon(\hat{x}) \subset \mathcal{G}^S(u)$ for all $u \in V_\delta(\hat{u})$. Thus, $\mathcal{G}^S$ is strongly lsc at $\hat{u}$ and the proof is complete. □

Apart from the cited statements from the book of Bank et al. [3], the following general result relating Hausdorff upper and lower semicontinuity for the singleton mappings is rather useful.

Lemma 2.35. *Let $\hat{u} \in \mathcal{U}$. If the mapping Γ is H-usc at $\hat{u}$ and the set $\Gamma(\hat{u})$ is a singleton, then Γ is also H-lsc (and hence B-lsc) at $\hat{u}$.*

Proof. Hausdorff upper semicontinuity guarantees that $\Gamma(u) \subset V_\varepsilon(\Gamma(\hat{u})), \forall \varepsilon > 0$, i.e. since $\Gamma(\hat{u})$ is a singleton, $\Gamma(u) \in V_\varepsilon(x(\hat{u}))$. Thus, for all $y \in \Gamma(u)$, the distance to the point $x(\hat{u})$ is less than ε, $d(y, x(\hat{u})) < \varepsilon$. Therefore, $x(\hat{u}) \in V_\varepsilon(y)$ for all $y \in \Gamma(u)$, i.e. $x(\hat{u}) \in V_\varepsilon(\Gamma(u))$. □

2.3.2 Properties of the feasibility set mapping $\mathcal{F}$

Naturally, the mapping $\mathcal{F}$ is the first one to start the investigations about continuity since this will be a crucial factor for the subsequent discussions of the other mappings.

Proposition 2.36. *The feasible set mapping $\mathcal{F}$ is closed and H-usc for all $u \in \mathcal{U}$.*

Proof. The set X is compact and the mapping $\mathcal{F}$ is closed since g is continuous and K is a closed cone, thus, $\mathcal{F}$ is Hausdorff upper semicontinuous according to Lemma 2.26 (i). □

Lemma 2.37. *If $g(\cdot, \hat{u})$ is strictly K-convex, then exactly one of the following is true:*

(i) The set $\mathcal{F}(\hat{u})$ contains only one element, i.e. $\mathcal{F}(\hat{u}) = \{x(\hat{u})\}$.

(ii) There exists a Slater point of $\mathcal{F}(\hat{u})$, i.e. $\mathcal{F}^S = \mathcal{G}^S(\hat{u}) \cap X \neq \emptyset$.

Proof. If $\mathcal{F}(\hat{u})$ contains only one element, we are done. So assume there are at least two elements within $\mathcal{F}(\hat{u})$, say x and y. Since $\mathcal{F}(\hat{u})$ is a convex set, the point $z := 0.5\,x + 0.5\,y$ also lies in $\mathcal{F}(\hat{u})$, and with $g(\cdot, \hat{u})$ strictly K-convex it follows that $g(z, \hat{u}) = g(0.5\,x + 0.5\,y, \hat{u}) <_K 0.5\,g(x, \hat{u}) + 0.5\,g(y, \hat{u}) \leq_K 0$ and thus z is a Slater point of $\mathcal{F}(\hat{u})$. □

The following proposition finally gives the first desired result, namely that the feasibility set mapping is Hausdorff continuous in case a Slater point exists. The proof relies on the extension of Theorem 3.1.5 in Bank et al. [3] to the conic setting which was presented in Theorem 2.34.

Proposition 2.38. *If there exists a Slater point of $\mathcal{F}(\hat{u})$, then $\mathcal{F}$ is H-lsc and hence H-continuous at $\hat{u}$.*

Proof. Let the mapping Γ be defined by $\Gamma := X$. Recall that using the definitions of $\mathcal{G}(u)$ and $\mathcal{G}^S(u)$, the set $\mathcal{F}(\hat{u})$ can be represented as $\mathcal{F}(\hat{u}) = \mathcal{G}(\hat{u}) \cap X$ and the associated set of Slater points is then $\mathcal{G}^S(\hat{u}) \cap X$ which is non-empty by assumption. We now apply Theorem 2.34 according to which it is sufficient to show that $(\mathcal{G} \cap \Gamma)(\hat{u}) \subset \mathrm{cl}(\mathcal{G}^S \cap \Gamma)(\hat{u})$, i.e. $(\mathcal{G} \cap X)(\hat{u}) \subset \mathrm{cl}(\mathcal{G}^S \cap X)(\hat{u})$. Using the same argument as in Proposition 2.28 this statement can be verified and thus Hausdorff lower semicontinuity is given by noting that $(\mathcal{G} \cap X)(\hat{u})$ is compact and applying Lemma 2.26, part (iii). Combining H-lsc with H-usc from Proposition 2.36 completes the proof. □

Proposition 2.39. *If $g(\cdot, \hat{u})$ is strictly K-convex, then $\mathcal{F}$ is H-lsc at $\hat{u}$.*

Proof. From Lemma 2.37 we have that either the set $\mathcal{F}(\hat{u})$ contains only one element or there exists a Slater point of $\mathcal{F}(\hat{u})$. In either case we are done by combining the results of Lemma 2.35 and Proposition 2.36 or by applying Proposition 2.38, respectively. □

In the previous two propositions we have shown that requiring one additional condition (either existence of a Slater point of $\mathcal{F}(\hat{u})$ or strict K-convexity of g) suffices to guarantee Hausdorff lower semicontinuity of $\mathcal{F}$. Together with the result of Proposition 2.36 that the mapping $\mathcal{F}$ is already Hausdorff upper semicontinuous under our general Assumptions 2.14, we have that $\mathcal{F}$ is Hausdorff continuous at $\hat{u}$.

Before investigating the other mappings, we shortly summarize all information (including the trivial ones) about the mapping $\mathcal{F}$:

- The set $\mathcal{F}(u)$ is compact and convex for all $u \in \mathcal{U}$.
- The mapping $\mathcal{F}$ is closed for all $u \in \mathcal{U}$.
- The mapping $\mathcal{F}$ is H-usc for all $u \in \mathcal{U}$.
- Requiring additionally either existence of a Slater point of $\mathcal{F}(u)$ for $u \in \mathcal{U}$ or strict K-convexity of $g(., u)$, then $\mathcal{F}$ is also H-lsc, thus Hausdorff continuous at u.

Having that under certain conditions the mapping $\mathcal{F}$ is continuous, we want to study in the following subsections the implications on the mappings $f^*, \mathcal{F}^*$ and $\mathcal{F}^*_\varepsilon$.

2.3.3 Properties of the optimal value function f^*

For the optimal value function f^* we obtain the following result.

Proposition 2.40. *Let $\mathcal{F}$ be Hausdorff continuous at $\hat{u} \in \mathcal{U}$. Then the extreme value function $f^* : \mathcal{U} \to \mathbb{R}$ is continuous at $\hat{u}$.*

Proof. This proof consists of the two parts of showing lower and upper semicontinuity of f^*.
Since the mapping $\mathcal{F}$ is Hausdorff continuous at $\hat{u}$ by assumption, it is H-usc at $\hat{u}$. Furthermore, we have that the objective function f is continuous, i.e. it is especially lower semicontinuous. With these prerequisites and $\mathcal{F}(\hat{u})$ being compact we can apply Theorem 2.31 (2) to deduce that f^* is lower semicontinuous at $\hat{u}$.
To prove upper semicontinuity of f^*, we use again the assumption that $\mathcal{F}$ is H-continuous at $\hat{u}$ which implies that $\mathcal{F}$ is B-lsc at $\hat{u}$. Together with the objective function f being upper semicontinuous we have the necessary conditions for applying Theorem 2.31 (1) and readily get that f^* is upper semicontinuous and therefore also continuous at $\hat{u}$. □

Combining the results from this Proposition 2.40 and Proposition 2.38 we have that the existence of a Slater point for $(GCP_{\hat{u}})$ is a sufficient condition for f^* being continuous in $\hat{u}$.

As we are mainly interested in only *continuity* of the optimal value function and hence have not explicitly considered any other features, we present a few selected results regarding the optimal value function that might be useful. For further insight we refer to Bank et al. [3] and especially Bonnans and Shapiro [18] who have analyzed extensively the properties of the optimal value function and the optimal set mapping in parametric optimization problems. They also study the link between the primal and the dual problem under various conditions and the consequences thereof for the set of optimal solutions of the dual program.

Proposition 2.41. *Let both $f \in \mathcal{C}^1(X \times \mathcal{U}, \mathbb{R})$ and $g \in \mathcal{C}^1(X \times \mathcal{U}, \mathbb{R}^m)$ and let $(GCP_{\hat{u}})$ possess a Slater point. Then*

(i) f^* is directionally differentiable at $\hat{u}$ in the direction d and

$$f^{*'}(\hat{u}; d) = \inf_{x \in \mathcal{F}^*(\hat{u})} \sup_{\lambda \in \Lambda(\hat{u})} (\nabla_u L(x, \lambda, \hat{u}))^T d$$

with $L(x, \lambda, \hat{u})$ denoting the Lagrangian of $(GCP_{\hat{u}})$ and $\Lambda(\hat{u})$ being the set of Lagrange multipliers of $(GCP_{\hat{u}})$,

(ii) under the additional assumptions of $\mathcal{F}^*(\hat{u}) = \{\hat{x}\}$ and $\Lambda(\hat{u}) = \{\hat{\lambda}\}$ being singletons, we furthermore get that f^* is differentiable at $\hat{u}$ and

$$\nabla f^*(\hat{u}) = \nabla_u L(\hat{x}, \hat{\lambda}, \hat{u}).$$

Remark 2.42.

(i) Note that we have

$$\nabla_u L(x, \lambda, \hat{u}) = \nabla_u f(x, \hat{u}) + \lambda^T \nabla_u g(x, \hat{u})$$

with $\nabla_u g(x, \hat{u})$ denoting the Jacobi matrix of g with respect to the variable u.

(ii) We would straightforwardly obtain directional differentiability and local Lipschitz continuity of f^* if f^* was convex, see e.g. Rockafellar [70], Theorem 23.4 and Theorem 10.4, respectively. But convexity of f^* is not generally given in our setting. Joint convexity[2] of $f(.,.)$ and $g(.,.)$ for all $(x, u) \in X \times \mathcal{U}$ would yield convexity of the extreme value function f^* as shown in the following. For ease of notation we first introduce the abbreviations

$$x_\alpha := \alpha x_1 + (1-\alpha)x_2,$$
$$u_\alpha := \alpha u_1 + (1-\alpha)u_2.$$

Consider

$$\begin{aligned} f^*(u_\alpha) &= f^*(\alpha u_1 + (1-\alpha)u_2) \\ &= \inf_{\substack{x_1, x_2 \in X \\ g(x_\alpha, u_\alpha) \leq_K 0}} f\big(\alpha x_1 + (1-\alpha)x_2, \alpha u_1 + (1-\alpha)u_2\big). \end{aligned}$$

Using joint convexity of f yields that

$$f^*(u_\alpha) \leq \inf_{\substack{x_1, x_2 \in X \\ g(x_\alpha, u_\alpha) \leq_K 0}} \alpha f(x_1, u_1) + (1-\alpha) f(x_2, u_2)$$

and as joint convexity of g implies that with the smaller feasibility set $\{g(x_1, u_1) \leq_K 0, g(x_2, u_2) \leq_K 0\}$ the function value at most increases, we can furthermore continue with the above being

$$\leq \inf_{\substack{x_1, x_2 \in X \\ g(x_1, u_1) \leq_K 0 \\ g(x_2, u_2) \leq_K 0}} \alpha f(x_1, u_1) + (1-\alpha) f(x_2, u_2)$$

[2]Joint convexity means that it holds for arbitrary $x_1, x_2 \in X$, $u_1, u_2 \in \mathcal{U}$ and $\alpha \in [0, 1]$ that

$$f(\alpha x_1 + (1-\alpha)x_2, \alpha u_1 + (1-\alpha)u_2) \leq \alpha f(x_1, u_1) + (1-\alpha) f(x_2, u_2).$$

which finally equals the weighted sum of the following two optimization problems, hence

$$\begin{aligned} &= \alpha \inf_{\substack{x_1 \in X \\ g(x_1,u_1) \leq_K 0}} f(x_1,u_1) + (1-\alpha) \inf_{\substack{x_2 \in X \\ g(x_2,u_2) \leq_K 0}} f(x_2,u_2) \\ &= \alpha f^*(u_1) + (1-\alpha) f^*(u_2). \end{aligned}$$

(iii) The optimal solution set $\mathcal{F}^*(\hat{u})$ is a singleton if e.g. the objective function f is strictly convex in x. There also exist conditions (some strict constraint qualifications) implying uniqueness of the Langrange multiplier $\hat{\lambda}$, see e.g. Proposition 4.47 in Bonnans and Shapiro [18].

Proof of Proposition 2.41.

(i) To show directional differentiability at $\hat{u}$, we consider the parametrization $\hat{u} + td \to \hat{u}$ with $t \downarrow 0$, i.e. $\varphi(t) = f^*(\hat{u} + td)$. As there exists a Slater point for $(GCP_{\hat{u}})$, i.e. for $t = 0$, we can apply Theorem 13 from [35] which gives existence of the onesided derivative $\varphi'_+(0)$ and also provides the following explicit formula:

$$\varphi'_+(0) = \inf_{x \in \mathcal{F}^*(\hat{u})} \sup_{\lambda \in \Lambda(\hat{u})} \frac{\partial}{\partial t} L(x, \lambda, \hat{u} + td)\Big|_{t=0}.$$

Expanding the gradient $\frac{\partial}{\partial t} L(x, \lambda, \hat{u} + td)$ according to the chain rule, we obtain

$$\varphi'_+(0) = f^{*'}(\hat{u}; d) = \inf_{x \in \mathcal{F}^*(\hat{u})} \sup_{\lambda \in \Lambda(\hat{u})} (\nabla_u L(x, \lambda, \hat{u}))^T d$$

and the result is proved.

(ii) If both the optimal solution $\hat{x}$ and the Langrange multiplier $\hat{\lambda}$ are unique, the formula in part (i) reduces to

$$f^{*'}(\hat{u}; d) = (\nabla_u L(\hat{x}, \hat{\lambda}, \hat{u}))^T d$$

which is linear in d and thus f^* is (Hadamard) differentiable according to Definition A.8. □

Further explicit results regarding differentiability of the optimal value function in specialized cases can e.g. be found in Bonnans, Shapiro [18], Chapter 4.3. The following example illustrates that convexity of f^* is not a necessary condition for f^* being directionally differentiable.

Example 2.43. *Consider the problem*

$$\min_{x\in[-1,1]} \quad ux$$

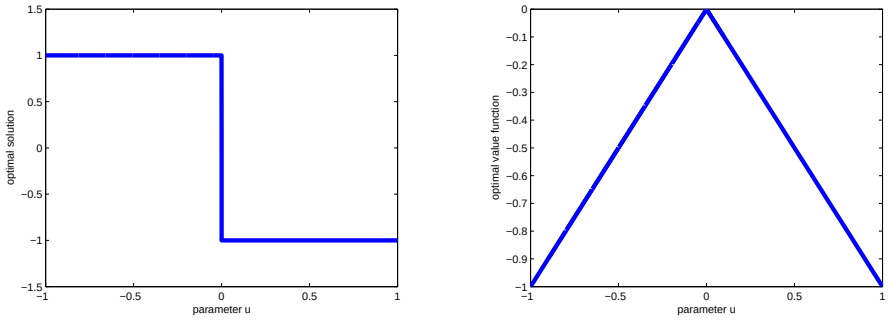
for $u \in [-1,1]$. All the general prerequisites are fulfilled. We especially have $f(x,u) = ux$ linear (thus convex) in u for fixed x, and in x for fixed u. Note that $f(x,u)$ is not jointly convex in (x,u). The optimal solution $\mathcal{F}^*(u)$ is given by

$$\mathcal{F}^*(u) = \begin{cases} \{1\} & \text{if } u < 0 \\ [-1,1] & \text{if } u = 0 \\ \{-1\} & \text{if } u > 0 \end{cases}$$

and the extreme value function can for all cases be expressed as

$$f^*(u) = -|u|.$$

It can be observed both in the formulas above and in Figures 2.3(a) and 2.3(b) that the optimal solution set $\mathcal{F}^*(u)$ is not a singleton for $u = 0$ and that f^* is continuous for all $u \in [-1,1]$. It is not differentiable at $u = 0$, but it is directionally differentiable for all $u \in [-1,1]$.



(a) optimal solution $\mathcal{F}^*(u)$ (b) optimal value function $f^*(u)$

Figure 2.3: Illustration of the optimal solution $\mathcal{F}^*$ and the extreme value function f^* in Example 2.43.

A further result stating pointwise Lipschitz continuity of the extreme value function is the following.

Proposition 2.44. *Suppose that $f \in \mathcal{C}^1(X \times \mathcal{U}, \mathbb{R})$ and $g \in \mathcal{C}^1(X \times \mathcal{U}, \mathbb{R})$, and let $(GCP_{\hat{u}})$ possess a Slater point. Then, the optimal value function f^* is pointwise Lipschitz continuous in $\hat{u}$.*

Proof. For pointwise Lipschitz continuity of f^* in $\hat{u}$ we need to show (Definition A.1) that there exists a neighborhood V around $\hat{u}$ and a constant $L = L(\hat{u}) > 0$ such that

$$\frac{\|f^*(u) - f^*(\hat{u})\|}{\|u - \hat{u}\|} \leq L \quad \forall u \in V.$$

Since f^* is directionally differentiable at $\hat{u}$ (see Proposition 2.41, part (i)) and X is finite dimensional, we can use that (see Proposition A.11 and Definition A.10)

$$f^*(\hat{u} + h) = f^*(\hat{u}) + f^{*'}(\hat{u}; h) + o(\|h\|).$$

With $V_{\tilde{\varepsilon}}(\hat{u})$ denoting a neighborhood around $\hat{u}$, consider the compact neighborhood $\overline{V_\varepsilon}(\hat{u})$ with $0 < \varepsilon < \tilde{\varepsilon}$ and let $h := u - \hat{u}$ with $u \in \overline{V_\varepsilon}(\hat{u})$, i.e. $h \in \mathbb{R}^d$ with $\|h\| \leq \varepsilon$. Then we have

$$\begin{aligned}
\frac{\|f^*(u) - f^*(\hat{u})\|}{\|u - \hat{u}\|} &= \frac{\|f^*(\hat{u} + h) - f^*(\hat{u})\|}{\|h\|} \\
&= \frac{\|f^{*'}(\hat{u}, h) + o(\|h\|)\|}{\|h\|} \\
&\leq \frac{\|f^{*'}(\hat{u}, h)\|}{\|h\|} + \frac{o(\|h\|)}{\|h\|} \\
&= \left\| f^{*'}\left(\hat{u}, \frac{h}{\|h\|}\right)\right\| + \frac{o(\|h\|)}{\|h\|} \\
&\quad \text{(as } f^{*'} \text{ is positively homogeneous in } h\text{, Lemma A.7)} \\
&\leq \sup_{\substack{h \in \mathbb{R}^d \\ \|h\| \leq \varepsilon}} \left\| f^{*'}\left(\hat{u}, \frac{h}{\|h\|}\right)\right\| + \frac{o(\|h\|)}{\|h\|}.
\end{aligned}$$

Since $f^{*'}$ is continuous in h (obvious from the representation in Proposition 2.41 (i)), the supremum over the compact set $\{h \in \mathbb{R}^d \mid \|h\| \leq \varepsilon\}$ is attained. Thus, the first expression is finite and bounded from above by $L_1 > 0$. Furthermore, the second expression tends to zero for $h \to 0$ and hence is bounded by $L_2 > 0$ for all $h \in \overline{V_\varepsilon}(0) = \{h \in \mathbb{R}^d \mid \|h\| \leq \varepsilon\}$ with ε sufficiently small. Therefore, we finally get

$$\frac{\|f^*(u) - f^*(\hat{u})\|}{\|u - \hat{u}\|} \leq L_1 + L_2 =: L$$

which proves pointwise Lipschitz continuity of f^* in $\hat{u}$. □

2.3.4 Properties of the optimal set mapping $\mathcal{F}^*$

Next we consider the optimal set mapping $\mathcal{F}^*$. First, we prove the special case of having a unique solution, which – together with existence of a Slater point – then

assures $\mathcal{U}$-stability of the program. Afterwards, we illustrate in small examples that both these assumptions cannot be weakened to achieve Hausdorff continuity of the optimal set mapping. In a more general setting without having a unique solution, only Hausdorff upper semicontinuity can be shown.

Theorem 2.45 ($\mathcal{U}$-stability). *Consider the program ($GCP_{\hat{u}}$) and*

(i) let the mapping $\mathcal{F}$ be Hausdorff continuous at $\hat{u} \in \mathcal{U}$,

(ii) assume that the program has a unique solution for $\hat{u}$.

Then, the optimal set mapping $\mathcal{F}^*$ is H-continuous at $\hat{u}$, i.e. ($GCP_{\hat{u}}$) is $\mathcal{U}$-stable.

Recall that the assumption of the program having only one single solution for the parameter $\hat{u}$ can for example be guaranteed if the objective function $f(x, \hat{u})$ is strictly convex in x for $\hat{u} \in \mathcal{U}$.

Proof. Having closedness of $\mathcal{F}$ and continuity of f^* (Proposition 2.40), applying Corollary 2.30 gives Berge and thus Hausdorff upper semicontinuity of $\mathcal{F}^*$ at $\hat{u}$. Applying Lemma 2.35 concludes the proof. □

Corollary 2.46. *Let $\mathcal{F}$ be continuous at u for all $u \in \mathcal{U}$ and let $\mathcal{F}^*(u)$ be a singleton for all $u \in \mathcal{U}$. Then the optimal mapping $\mathcal{F}^*$ is Hausdorff continuous on $\mathcal{U}$, i.e. the family (GCP_u) is $\mathcal{U}$-stable.*

Proof. Follows directly from Theorem 2.45. □

Theorem 2.45, or Corollary 2.46 respectively, assures the desired result of having a single-valued continuous optimal solution. But there are two relatively strong requirements: the strict convexity of the objective function – or respectively, it is enough to assume uniqueness of the optimal solution – and the Hausdorff continuity of the feasible set mapping $\mathcal{F}$. In the following small examples we want to illustrate that both conditions of Theorem 2.45 (i.e. Hausdorff continuity of $\mathcal{F}$ and a unique solution) are necessary, neglecting or weakening only one of them does not suffice to give the desired result in general.

Example 2.47. *In this simple example we demonstrate that only convexity (and not strict convexity which would yield a unique solution) of the objective function is not sufficient to assure continuity in the solution.*

Consider the program

$$\min_{x \in [-1,1]} \quad ux$$

Let $\mathcal{U}$ be the compact interval $[-1, 1]$. In this case the set of feasible points is the same for each $u \in \mathcal{U}$, $\mathcal{F}(u) = \{x \in \mathbb{R} \mid -1 \leq x \leq 1\}$ and therefore $\mathcal{F}$ is

continuous (since constant). The objective function is linear, thus convex, but the optimal solution is

$$\mathcal{F}^*(u) = \begin{cases} 1 & \text{if } u < 0 \\ [-1,1] & \text{if } u = 0 \\ -1 & \text{if } u > 0 \end{cases}$$

and thus neither H- nor B-continuous.

Figure 2.4 illustrates the set of feasible points and the associated optimal solution for each $u \in \mathcal{U} = [-1,1]$ of the above program.

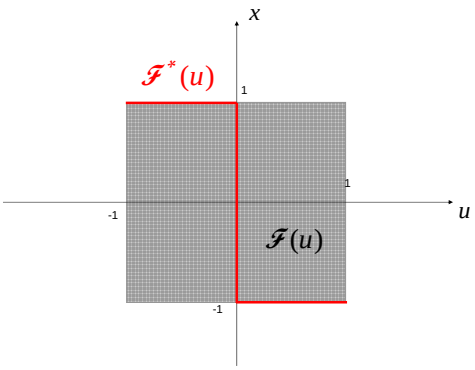


Figure 2.4: Illustration of the feasibility set and the associated optimal solution in Example 2.47.

Example 2.48. *This second example demonstrates the necessity of the second crucial requirement of $\mathcal{F}$ being continuous. Consider the program*

$$\begin{aligned} \min_{x \in [-1,1]} \quad & (x-1)^2 \\ s.t. \quad & ux \leq 0. \end{aligned}$$

again with $\mathcal{U} = [-1,1]$. In this program, the objective function is independent of the parameter u and it is strictly convex. The set of feasible points is given by

$$\mathcal{F}(u) = \begin{cases} [0,1] & \text{if } u < 0 \\ [-1,1] & \text{if } u = 0 \\ [-1,0] & \text{if } u > 0 \end{cases}$$

and therefore not continuous. And as well, the optimal solution

$$\mathcal{F}^*(u) = \begin{cases} 1 & \text{if } u < 0 \\ 1 & \text{if } u = 0 \\ 0 & \text{if } u > 0 \end{cases}$$

is not continuous in u. Note that for the particular parameter $u = 0$, no Slater point can be found. This already leads to the conjecture that continuity might not be achieved in that point – which is the case as shown by the explicit results.

Figure 2.5 illustrates the set of feasible points and the associated optimal solution for each $u \in \mathcal{U} = [-1, 1]$ of the above program.

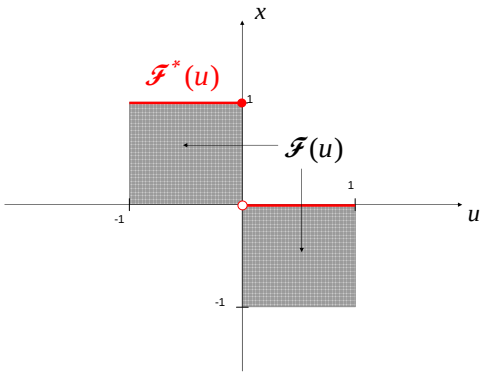


Figure 2.5: Illustration of the feasibility set and the associated optimal solution in Example 2.48.

After having examined the special case of $\mathcal{F}^*$ being Hausdorff continuous if $\mathcal{F}^*(u)$ is a singleton, we explore the general optimal set mapping $\mathcal{F}^*$ with respect to continuity results.

Proposition 2.49. *Let $\mathcal{F}$ be Hausdorff continuous at $\hat{u} \in \mathcal{U}$. Then, the optimal set mapping $\mathcal{F}^*$ is closed at $\hat{u}$.*

Proof. Follows directly from Theorem 2.29 since continuity of $\mathcal{F}$ at $\hat{u}$ implies continuity of f^* at $\hat{u}$. □

Proposition 2.50. *Let $\mathcal{F}$ be Hausdorff continuous at $\hat{u} \in \mathcal{U}$. Then, the optimal set mapping $\mathcal{F}^*$ is Hausdorff upper semicontinuous at $\hat{u}$.*

Proof. Having closedness of $\mathcal{F}$ and continuity of f^* (Proposition 2.40), applying Corollary 2.30 gives Berge and thus Hausdorff upper semicontinuity of $\mathcal{F}^*$ at $\hat{u}$. □

Thus, in the general case of possibly having an entire set of optimal solutions, we can only guarantee Hausdorff upper semicontinuity, even in rather simple optimization problems, as seen in Example 2.47 where $\mathcal{F}$ was constant and $f(x, u)$ was linear both in x and u. To achieve H-continuity of $\mathcal{F}^*$ (i.e. Hausdorff lower semicontinuity), some rather strong conditions on (GCP_u) are required, see Bonnans and Shapiro [16].

2.3.5 Properties of the ε-optimal set mapping $\mathcal{F}_\varepsilon^*$

So far we have established the main results concerning continuity of the feasibility set mapping, the extreme value function and the mapping onto the set of optimal solutions. As only Hausdorff upper semicontinuity could be achieved for the optimal set mapping, we now investigate the mapping onto the set of ε-optimal solutions with respect to continuity. We will find that the mapping $\mathcal{F}_\varepsilon^*$ possesses the property of being Hausdorff lower semicontinuous which at least assures the existence of as continuous selection function, see page 23. Closedness additionally yields Hausdorff upper semicontinuity, thus Hausdorff continuity. These results are summarized and proved in the following proposition.

Proposition 2.51. *Let $\mathcal{F}$ be Hausdorff continuous at $\hat{u} \in \mathcal{U}$. Then the ε-optimal set mapping $\mathcal{F}_\varepsilon^*$ is H-continuous at $\hat{u}$ for all $\varepsilon > 0$.*

Before being able to prove this proposition, we need to cite a further result from Bank et al. [3]:

Lemma 2.52. *Let the mapping $\Gamma_1 : \mathcal{U} \to \mathcal{P}(X)$ be B-lsc at $\hat{u}$ and $\Gamma_2 : \mathcal{U} \to \mathcal{P}(X)$ be strongly lsc at $\hat{u}$. Then, the mappings*

$$(\Gamma_1 \cap \Gamma_2)(u) := \Gamma_1(u) \cap \Gamma_2(u)$$

and

$$\mathrm{cl}(\Gamma_1 \cap \Gamma_2)(u) := \mathrm{cl}(\Gamma_1(u) \cap \Gamma_2(u))$$

are B-lsc at $\hat{u}$.

Proof. See Bank et al. [3], Lemma 2.2.5. □

Proof of Proposition 2.51. For proving that the mapping $\mathcal{F}_\varepsilon^*$ is H-continuous, we need to show Hausdorff upper and lower semicontinuity.

- For Hausdorff upper semicontinuity it suffices to prove closedness of the mapping $\mathcal{F}_\varepsilon^*$ according to Lemma 2.26 (i). Let $\{u_k\} \subset \mathcal{U}$ be a sequence with $u_k \to \hat{u}$, and let $x_k \in \mathcal{F}_\varepsilon^*(u_k), x_k \to \hat{x}$. Since the mapping $\mathcal{F}$ is closed at $\hat{u}$, $\hat{x} \in \mathcal{F}(\hat{u})$. Having $x_k \in \mathcal{F}_\varepsilon^*(u_k)$ gives $f(x_k, u_k) - f^*(u_k) \leq \varepsilon$ and since both f and f^* are continuous, we get $f(\hat{x}, \hat{u}) - f^*(\hat{u}) \leq \varepsilon$. Thus, $\hat{x} \in \mathcal{F}_\varepsilon^*(\hat{u})$.

- To prove Hausdorff lower semicontinuity we use $\widetilde{\mathcal{F}}_\varepsilon^* = \widetilde{\mathcal{F}}_\varepsilon^* \cap \mathbb{R}^n$ with

  $$\widetilde{\mathcal{F}}_\varepsilon^* = \{x \in \mathcal{F}(u) \mid f(x,u) < f^*(u) + \varepsilon\}$$

  being B-lsc at $\hat{u}$ (Propositions 2.40 and Corollary 2.32) and the mapping Γ with $\Gamma(u) = \mathbb{R}^n$ being strongly lower semicontinuous. Thus, Lemma 2.52 yields B-lsc of $\mathcal{F}_\varepsilon^* = \mathrm{cl}(\widetilde{\mathcal{F}}_\varepsilon^* \cap \mathbb{R}^n)$ at $\hat{u}$. H-lsc follows from compactness of $\mathcal{F}_\varepsilon^*(\hat{u})$ and Lemma 2.26 (iii). □

2.3.6 Illustrative example

In the following example we will investigate a simple 2-dimensional optimization problem with a linear objective function and without any parametric constraints. In this case it is possible to explicitly calculate the optimal solution and also present some formulas for ε-optimal solutions.

Consider the following program[3]:

$$\min_{x \in X} \quad -x^T u \tag{P}$$

with $X = \{x \in \mathbb{R}^2 \mid x \geq 0,\ x^T\mathbf{1} = 1\}$, thus compact, and $u \in \mathcal{U}$. The notation "$\mathbf{1}$" stands for the appropriately sized vector consisting of 1 in each component. Before digging into detailed analysis of the program, we want to summarize some immediate statements about this particular problem and its characteristics concerning stability:

- If the two components of a parameter realization $\hat{u}$ are equal, i.e. w.l.o.g. $\hat{u} = \hat{u}_1 \cdot \mathbf{1}$, the set of optimal solutions for this particular $\hat{u}$ is not a singleton, but the entire feasibility set, i.e. $\mathcal{F}^*(\hat{u}) = \mathcal{F}(\hat{u}) = X$, since

  $$\min_{x \in X} \ -x^T\hat{u} = \min_{x \in X} \ -\hat{u}_1(\underbrace{x^T\mathbf{1}}_{=1}) = \min_{x \in X} \ -\hat{u}_1 = -\hat{u}_1.$$

  Thus, it is doubtable that the optimal set mapping is continuous at such an $\hat{u} = \hat{u}_1 \cdot \mathbf{1} \in \mathcal{U}$.

- Since there are no constraints depending on u, the feasible set mapping $\mathcal{F}$ is constant and thus H-continuous on $\mathcal{U}$. Hence, we know that there exist at least continuous selection functions within the set of ε-optimal solutions $\mathcal{F}^*_\varepsilon$, see page 23.

- From $\mathcal{F}$ being H-continuous on $\mathcal{U}$, we can further deduce that the extrem value function f^* is continuous, thus the optimal value will not change much even though the optimal solution (i.e. $\mathcal{F}^*(u)$) itself can be quite different if the parameter u is disturbed only slightly.

In this simple example we can explicitly state the extreme value function $f^*(u)$ and the optimal set mappings $\mathcal{F}^*(u)$ and $\mathcal{F}^*_\varepsilon$:

- The extreme value function for any $u \in \mathcal{U}$ is given by

  $$f^*(u) = \min_{x \in X} \ -x^T u = -\max\{u_1, u_2\}.$$

[3]This problem is rather well-known in asset management as it represents maximizing the portfolio return. This and related problems will be introduced and studied in Chapters 4 to 7.

- The associated optimal set mapping $\mathcal{F}^*$ is given by

$$\mathcal{F}^*(u) = \{x \in X \mid -x^T u = f^*(u)\}$$
$$= \begin{cases} \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} & \text{if } u_1 > u_2 \\ \left\{ \lambda \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (1-\lambda) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mid 0 \le \lambda \le 1 \right\} = X & \text{if } u_1 = u_2 \\ \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} & \text{if } u_1 < u_2 \end{cases}$$

which is obviously neither B- nor H-lsc at any u with $u_1 = u_2$, confirming our previously stated expectation. Furthermore, it is impossible to find a continuous selection function $\zeta^* : \mathcal{U} \to X, u \to x^*(u)$ within the mapping $\mathcal{F}^*$. Moreover, as already anticipated in the theoretical considerations above, the extreme value function f^* is continuous at all $u \in \mathcal{U}$.

Figure 2.6 illustrates the optimal solution (represented here by the first component, the second is then determined as well by the constraint $x^T\mathbf{1} = 1$) for all possible parameter vectors $u \in \mathcal{U}$. In view of extending this example later on where we will investigate the influence of different shapes of the uncertainty set $\mathcal{U}$, we plot $\mathcal{F}^*(u)$ on the basis of $\mathcal{U}$ once chosen as a square and once as a circle. It can be nicely observed that $\mathcal{F}^*$ is discontinuous at the line of parameters u with $u_1 = u_2$.

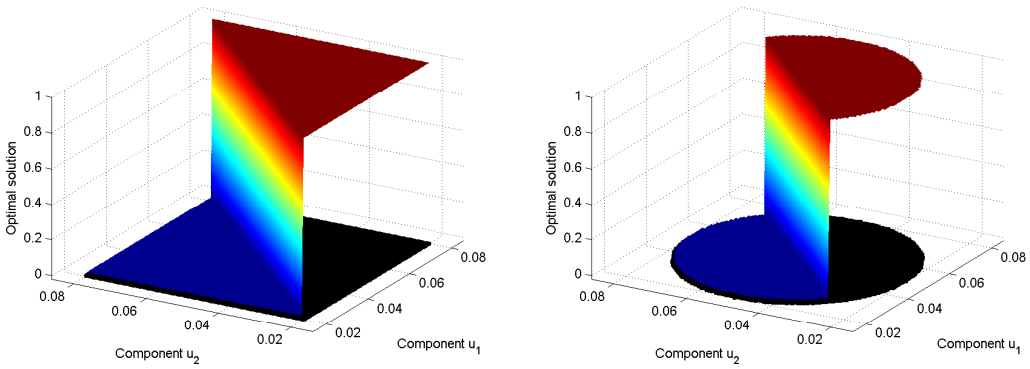


Figure 2.6: Illustration of the optimal solution for parameters $u \in \mathcal{U}$, using two different shapes of $\mathcal{U}$.

After having confirmed the expectations for f^* and $\mathcal{F}^*$, we now examine explicitly the ε-optimal set mapping.

- The ε-optimal set mapping for $\varepsilon > 0$ is given in terms of

$$\mathcal{F}_\varepsilon^*(u) = \{x \in X \mid -x^T u \leq f^*(u) + \varepsilon\}$$
$$= \begin{cases} \left\{ \begin{pmatrix} \lambda \\ 1-\lambda \end{pmatrix} \mid \max\left(0; 1 - \frac{\varepsilon}{u_1 - u_2}\right) \leq \lambda \leq 1 \right\} & \text{if } u_1 > u_2 \\ \left\{ \begin{pmatrix} \lambda \\ 1-\lambda \end{pmatrix} \mid 0 \leq \lambda \leq 1 \right\} & \text{if } u_1 = u_2 \\ \left\{ \begin{pmatrix} \lambda \\ 1-\lambda \end{pmatrix} \mid 0 \leq \lambda \leq \max\left(\frac{\varepsilon}{u_2 - u_1}; 1\right) \right\} & \text{if } u_1 < u_2. \end{cases}$$

Even though it was impossible to define a continuous selection function consisting only of truly optimal solutions, we can find one within the set of ε-optimal solutions, since the ε-optimal set mapping $\mathcal{F}_\varepsilon^*$ allows an entire range of feasible points in the cases where $u_1 \neq u_2$. A working definition of a continuous selection function for feasible $u \in \mathcal{U}$ is e.g. given by

$$\zeta_\varepsilon^*(u) := \begin{pmatrix} \lambda \\ 1-\lambda \end{pmatrix} \quad \text{with} \quad \lambda := \begin{cases} 1 - \frac{1}{2}\frac{\varepsilon}{|u_2 - u_1|} & \text{if } u_2 - u_1 < -\varepsilon \\ \frac{1}{2} & \text{if } -\varepsilon < u_2 - u_1 < \varepsilon \\ \frac{1}{2}\frac{\varepsilon}{|u_2 - u_1|} & \text{if } u_2 - u_1 > \varepsilon. \end{cases}$$

This definition of a selection function only depends on the difference between u_1 and u_2 and its relation to ε. Figure 2.7 illustrates the range of possible portfolio allocations (characterized here by the relative weight in the first asset) for $\varepsilon = 0.1$. The black line denotes the truly optimal solutions (i.e. solutions that lie within the optimal set mapping $\mathcal{F}^*(u)$) which flip from $(0,1)^T$ to $(1,0)^T$ at the point where the difference $u_2 - u_1$ becomes zero. The blue line indicates the relaxed bound on the components of x if ε-optimal solutions suffice, thus a whole range is possible in this case. The red line shows the above choice of selection function lying within the region of ε-optimal solutions.

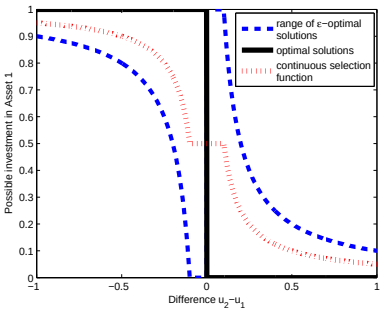


Figure 2.7: Continuous selection function within $\mathcal{F}_\varepsilon^*$.

2.3.7 Summary

To round off this section about $\mathcal{U}$-stability, we want to collect the main results: We first of all found that the feasible set mapping $\mathcal{F}$ is always Hausdorff upper continuous in our general setting. Hausdorff continuity (thus basically H-lsc) can be guaranteed if additionally the constraint qualification that a Slater point exists is satisfied. These results lay the foundation for the analysis of the (ε-)optimal set mappings and the extreme value function.

Assuming Hausdorff continuity of the feasible set mapping $\mathcal{F}$, Corollary 2.46 then assures (H-)continuity of the optimal solution in the case that there is only one unique optimal solution for each parameter $u \in \mathcal{U}$. As in the general multi-valued case the mapping $\mathcal{F}^*$ is not necessarily continuous, we extend the scope of the study to the ε-optimal set mapping $\mathcal{F}^*_\varepsilon$. Both these mappings can be proved to be Hausdorff lower semicontinuous which – according to page 23 – is sufficient for the existence of a continuous selection function. Closedness of $\mathcal{F}^*_\varepsilon$ furthermore yields H-usc and thus Hausdorff continuity of $\mathcal{F}^*_\varepsilon$.
Moreover, Hausdorff continuity of $\mathcal{F}$ suffices to show continuity of the optimal value function f^*.

Remark 2.53. *Regarding local Lipschitz continuity of f^* and $\mathcal{F}^*$, the following selected results give more insights into necessary and sufficient conditions. For more details and results, we refer e.g. to Bonnans and Shapiro [16, 17, 18] who have extensively studied optimization problems with respect to Lipschitz continuity.*

- *Under the condition that $\mathcal{F}$ is independent of u and thus constant, we have the subsequent statements. Recall that strict convexity of f in x and hence uniqueness of the optimal solution was sufficient for $\mathcal{F}^*$ being continuous (see Theorem 2.45), but it is not sufficient for $\mathcal{F}^*$ being locally Lipschitz continuous. A counterexample is given in Bonnans and Shapiro [16], Example 6.1.*

- *In the more general situation of $\mathcal{F}$ being dependent on u, there exist conditions that (together with existence of a Slater point) guarantee local Lipschitz continuity of $\mathcal{F}^*$ at $\hat{u}$ (see Bonnans, Cominetti and Shapiro [17], Theorem 3.1).*

- *In the special case of u representing canonical perturbations, i.e. $f(x,u) = f(x) + x^T u$ and $g(x,u) = g(x) + u$, Dontchev [24] gives conditions such that the optimal solution is locally Lipschitz continuous.*

Finally, Figure 2.8 graphically illustrates all the results and implications we have established in this section.

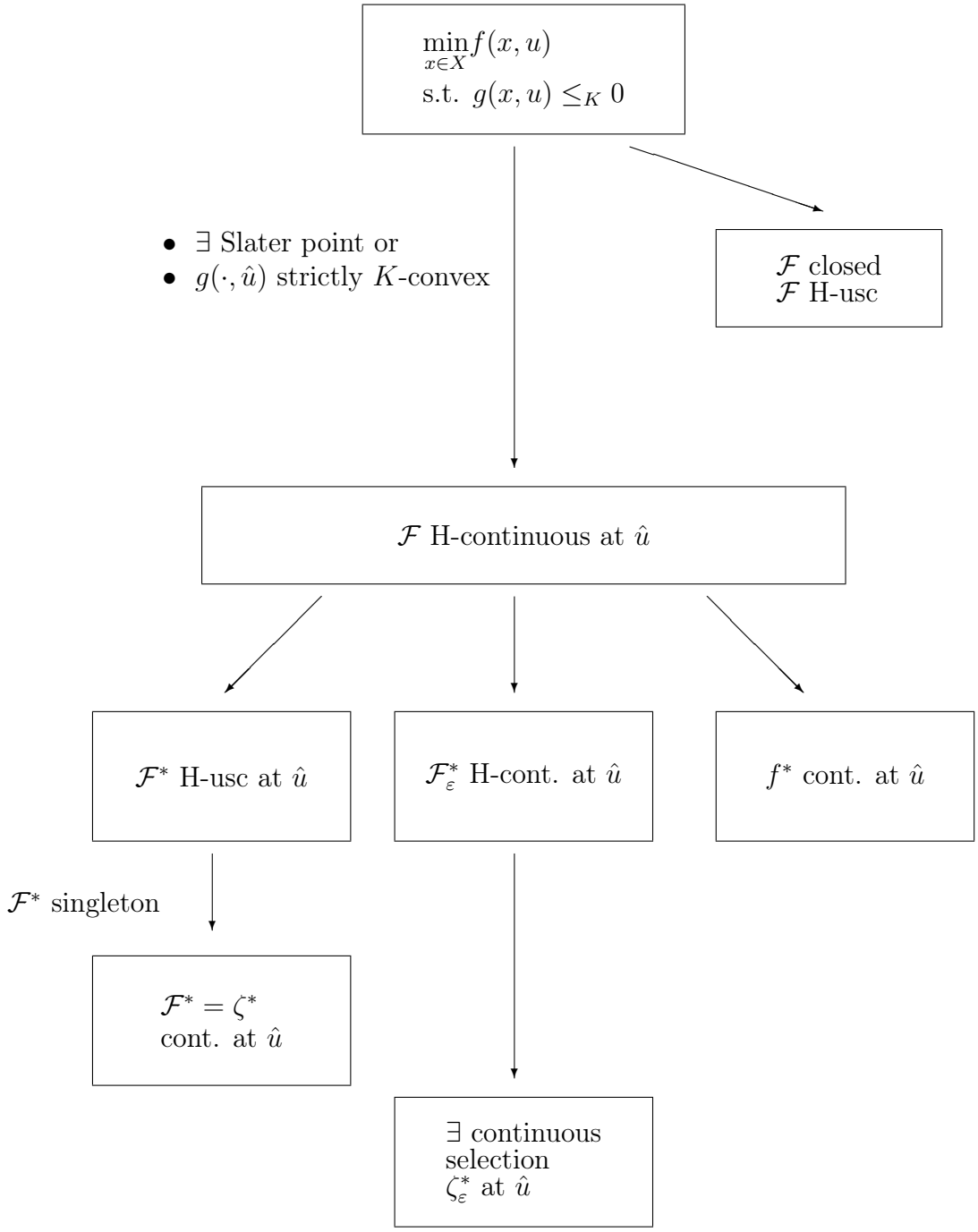


Figure 2.8: Illustration of the continuity results of Section 2.3.

Chapter 3

The (local) robust counterpart approach

In this chapter we present and discuss the so-called *robust counterpart approach* as introduced in 1998 by Ben-Tal and Nemirovski and discussed in various papers, see for example [4, 5, 6, 7, 8, 9]. The same approach was independently presented by El-Ghaoui, Oustry and Lebret [26]. It relies on the basic idea to use not only a particular point estimate instead of the uncertain parameter to solve the optimization problem, but to consider an entire set of possible parameter realizations, an *uncertainty set*. The robust counterpart approach is a worst case approach, as the optimization problem is now solved under the constraint that any point within the uncertainty set could be chosen, i.e. especially the one leading to the worst performance.

Since such a robustification (usually) changes the type of the optimization to a more difficult class[1], see e.g. [8], there are cases where the robust problem cannot be solved anymore with standard techniques. Hence, tractable approximations to the robust formulation have to be found, see e.g. Ben-Tal and Nemirovski [4, 8] and Bertsimas and Sim [12]. Ben-Tal, Boyd and Nemirovski extended the idea of the robust counterpart in [10] to additional treatment of data realizations lying outside the considered uncertainty set. The recent paper "Selected topics in robust convex optimization" of Ben-Tal and Nemirovski [11] nicely overviews the robust counterpart approach including the extended idea and tractability. It furthermore contains a comprehensive list of references to literature on robust optimization.

In this Chapter we first introduce the necessary definitions and notation and present the local robust counterpart approach. Afterwards, we investigate the continuity characteristics of the robust optimization program analogously to Section 2.3. Furthermore, we examine under which limitations this approach works

[1] Roughly speaking, it holds that a linear program (LP) becomes a second-order cone program (SOCP), a SOCP becomes a semidefinite program (SDP) and an SDP results in a problem that is usually not solvable in polynomial time, see [4].

and what consequences result thereof.

To our knowledge there do not exist any explicit stability investigations of the robust counterpart program. We study the respective continuity characteristics in Section 3.2. Foreclosing some results, it can be stated that none of the nice continuity properties are lost when modifying the original problem to its robust counterpart. Thus, we again obtain Hausdorff upper semicontinuity of the optimal set mapping in case there exists a Slater point for the robust program. And furthermore, it still holds that if the solution is unique, the optimal set mapping is continuous.

Additionally, it holds that the robust solution converges to the solution of the original problem if the uncertainty set shrinks to a single point, i.e. it then merely is a point estimate of the unknown parameter. Such a behavior is both desired and intuitively expected.

As far as we are aware of, there are no studies about the influence of the particular *shape* of the chosen uncertainty set on the optimal solution. We investigate in Section 3.3 how the shape affects the optimization result. An example illustrates that interval or box uncertainty does not guarantee any advantage with respect to continuity or uniqueness of the optimal solution, whereas it can be shown that under rather general assumptions an ellipsoidal uncertainty set yields a particular structure of the set of optimal solutions. An important consequence thereof is that in most practical cases (e.g. in virtually all portfolio optimization problems in asset management) the optimal solution is unique and thus continuous. Hence, this illustrates that robustification using an ellipsoidal uncertainty set can achieve $\mathcal{U}$-stability of the problem.

Since such a benefit of having a unique and continuous optimal solution cannot be obtained for free, we investigate in Section 3.4 the costs associated with this approach, where costs are represented by an increase of the optimal objective value. Naturally, the solution of the robust counterpart problem crucially depends on the *size* of the chosen uncertainty set. It is shown in Section 3.4 that the increase in objective value is linear in the size δ of the uncertainty set for small δ. The same result was given by Ben-Tal and Nemirovski in [4], but only for linear optimization problems and under different conditions. We prove this statement for our generalized convex conic optimization problem.

The question in practical applications is how to define the particular uncertainty set for the problem at hand. Choosing it too large might result in an empty feasibility set or yields an optimal solution which is too conservative to be of practical use. On the other hand, defining the uncertainty set too small does not account for the possible variations of the unknown parameter we intended to capture.

In the existing literature about application of the robust counterpart approach there are many approaches to define uncertainty sets, most of them being interval or box uncertainty sets or ellipsoids – or mixtures of both. Ben-Tal and Nemirovski [4, 5] and Ben-Tal, El-Ghaoui and Nemirovski [6] propose ellipsoids

or intersections of ellipsoids as uncertainty sets, which are used as well in Lutgens [55]. On the other hand, Tütüncü and Koenig [79] and also El-Ghaoui, Oks and Oustry [27] use intervals to characterize uncertainty, and Goldfarb and Iyengar [33] use both intervals and ellipsoids to define uncertainty sets around the parameters of their factor model used to describe the asset return distribution. We address the problem of the *shape* of the uncertainty set in Section 3.3 from a theoretical point of view, and in Chapter 5 we discuss explicit definitions of uncertainty sets for the portfolio optimization problem from asset management.

3.1 General definitions

This section contains the introduction of the idea of the robust counterpart approach according to Ben-Tal and Nemirovski [4, 8] and the associated definitions. We furthermore extend their concept to a *localized* version.

Definition 3.1. *A point $x \in X$ is called* worst-case-$\mathcal{U}$-robust *if and only if $x \in \mathcal{F}_{\mathcal{U}}$, i.e. if*

$$g(x,u) \leq_K 0 \quad \textit{for all } u \in \mathcal{U}.$$

This Definition 3.1 is a definition of *worst-case-robustness* of feasible points of (GCP_u), meaning that a candidate solution x is worst-case-$\mathcal{U}$-robust if it is a feasible point *for all* $u \in \mathcal{U}$, i.e. no matter which possible parameter realization is considered.

Definition 3.2. *The* robust counterpart *to the family (GCP_u), $u \in \mathcal{U}$ is given by the semi-infinite program*

$$\begin{aligned} \min_{x \in X} \quad & \max_{u \in \mathcal{U}} f(x,u) \\ s.t. \quad & g(x,u) \leq_K 0 \quad \forall u \in \mathcal{U}. \end{aligned} \tag{RC}$$

Note that as we have assumed $\mathcal{F}_{\mathcal{U}}$ to be non-empty (Assumption 2.17), this guarantees that the feasibility set of the robust counterpart program is non-empty as well.

Remark 3.3. *The optimization problem (RC) can equivalently be expressed as*

$$\min_{x \in \mathcal{F}_{\mathcal{U}}} \quad \max_{u \in \mathcal{U}} f(x,u).$$

For applying the classical saddle point theory to this problem we would need linearity of f in u (or more generally, concavity in u) which is not assumed in our setting. In their paper "Robust asset allocation", Tütüncü and Koenig [79] make use of a saddle point approach for solving a robust counterpart problem in finance.

Remark 3.4. *When solving the generic convex program (GCP_u), it is a well known fact that we can without loss of generality*[2] *assume the objective function to be independent of the uncertain parameter u and linear in x. This will – whenever the assumption is convenient – be denoted by $f(x,u) = l(x)$. Such a simplifying assumption can as well be made without loss of generality when dealing with the robust counterpart program. For a more detailed discussion see Appendix F.*

We will make use of this simplification for the robust program when proving Theorem 3.37.

For actual programming we need to deal with the expression "for all parameters $u \in \mathcal{U}$". If the uncertainty set $\mathcal{U}$ is a finite set, we can simply replace the one constraint by finitely many constraints for each single parameter $u \in \mathcal{U}$. Analogously, if the uncertainty set has a finite number of vertices – e.g. when using the convex hull of some points (finitely many) as an uncertainty set – it suffices to consider the contraint function only at these vertices. But if $\mathcal{U}$ is not finite, we are in the field of semi-infinite programming (SIP). In our practical problems the semi-infinite constraint can be reformulated by determining and inserting the worst case parameter of the uncertainty set. If such a transformation to a classical convex optimization problem is not possible, the solution of the semi-infinite program is usually approximated iteratively by solving the problem with increasing but finitely many constraints. For more details and literature about semi-infinite programming, we refer to the books of Goberna and López [32] and Reemtsen and Rückmann [67].

In our setting, to apply the robust counterpart approach we need to have a precise definition of the uncertainty set $\mathcal{U}$ in the actual application. To create a particular uncertainty set around a point estimate, there are two further necessary characteristics – the size and the shape.

The question of the shape of the uncertainty set is examined in Subsection 3.3, where we will consider the robust counterpart using the two most intuitive shapes, interval or box uncertainty and ellipsoidal uncertainty.

To address the problem of the size we define a *local robust counterpart* where the constraint gets relaxed in that it does not have to hold for all $u \in \mathcal{U}$ but only for those u within a smaller region around a certain parameter choice $\hat{u}$. Thus this smaller uncertainty set represents some kind of local robustness. The effect of diminishing the uncertainty set will be investigated in Subsection 3.4.

Before defining explicitly the local robust counterpart program (which is nothing else than the robust counterpart as given in Definition 3.2 using a smaller uncertainty set), we introduce the notion of a local uncertainty set.

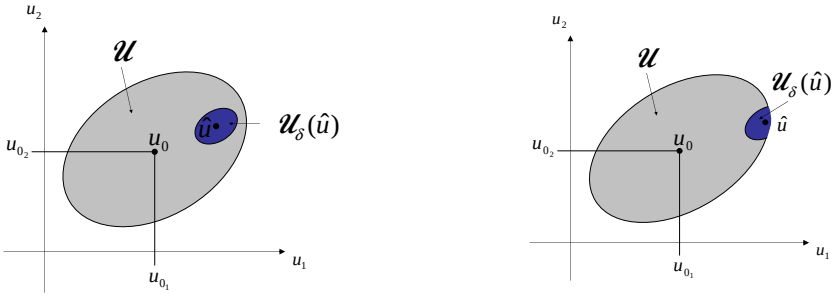
Notation 3.5. *If not explicitely stated otherwise, the uncertainty set $\mathcal{U}$ is supposed to be "centered" at u_0, i.e. we can write $\mathcal{U} = u_0 + \mathcal{U}'$ with $\mathcal{U}'$ such that*

[2] We can always introduce an additional variable to be minimized and move the objective function to the set of constraints.

$0 \in \mathcal{U}'$. *An advantage of this representation of* $\mathcal{U}$ *is that the size of the uncertainty part* $\mathcal{U}'$ *is now scalable. Since we will introduce local robustness in the following, we define a smaller uncertainty set around some point* $\hat{u} \in \mathcal{U}$ *and with suitably chosen size* $\delta \geq 0$ *by*

$$\mathcal{U}_\delta(\hat{u}) = \hat{u} + \delta\mathcal{U}' \cap \mathcal{U}$$

with $\delta\mathcal{U}' = \{\delta v \mid v \in \mathcal{U}'\}$. *Figure 3.1 illustrates the relation of* $\mathcal{U}$ *and* $\mathcal{U}_\delta(\hat{u})$ *and the introduced notational convention. Figure 3.1(a) shows the case where* $\hat{u} + \delta\mathcal{U}' \subset \operatorname{int}\mathcal{U}$ *and Figure 3.1(b) describes the case where an intersection of* $\hat{u} + \delta\mathcal{U}'$ *with* $\mathcal{U}$ *is needed to restrict possible parameter realization to the given larger set* $\mathcal{U}$.



(a) $\mathcal{U}_\delta(\hat{u})$ within $\operatorname{int}\mathcal{U}$ (b) $\mathcal{U}_\delta(\hat{u})$ as intersection with $\mathcal{U}$

Figure 3.1: Illustration of $\mathcal{U}$ and $\mathcal{U}_\delta(\hat{u})$ in the two-dimensional case with an ellipsoidal shape.

Remark 3.6. *As we have assumed the large uncertainty set* $\mathcal{U}$ *to be non-empty, convex and compact (Assumption 2.14), the local uncertainty set* $\mathcal{U}_\delta(\hat{u})$ *is obviously non-empty, convex and compact as well.*

For an easier characterization of "suitably chosen" midpoint $\hat{u}$ and size δ we make the following definition:

Definition 3.7. *We call a pair* $(\hat{u}, \delta)$ admissible *if and only if* $\hat{u} \in \mathcal{U}$ *and* $\delta \in \mathbb{R}_+$ *such that* $\hat{u} + \delta\mathcal{U}' \subset \mathcal{U}$, *i.e. the intersection with* $\mathcal{U}$ *is not necessary.*

Figure 3.1(a) illustrates the case of $(\hat{u}, \delta)$ being admissible. In the case of $(\hat{u}, \delta)$ being admissible the local uncertainty set maintains its originally chosen shape, e.g. an ellipse, and does not include any artificially introduced vertices.

In the following, if not explicitly stated otherwise, the considered pairs $(\hat{u}, \delta)$ are admissible.

Having all the necessary definitions and notation, we can state the appropriate optimization problem. The associated program to actually find a locally worst-case $\mathcal{U}$-robust solution will be called *local robust counterpart* (LRC) and is defined as follows.

Definition 3.8. *Let $\hat{u} \in \mathcal{U}$ and $\delta \geq 0$. Then the local robust counterpart ($LRC_{\hat{u},\delta}$) to the program ($GCP_{\hat{u}}$) is*

$$\min_{x \in \mathcal{F}_{\mathcal{U}_\delta(\hat{u})}} \max_{u \in \mathcal{U}_\delta(\hat{u})} f(x,u) \quad = \quad \min_{x \in X} \max_{u \in \mathcal{U}_\delta(\hat{u})} f(x,u) \qquad (LRC_{\hat{u},\delta})$$
$$g(x,u) \leq_K 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}).$$

Note that this is the same as Definition 3.2 except for the smaller (i.e. local) uncertainty set. The difference hence is that in the program ($LRC_{\hat{u},\delta}$) the robustifying condition has to be fulfilled only for some of the possible parameter realizations u, but not for all $u \in \mathcal{U}$. According to Definition 3.1 the optimal solution of the local robust counterpart program (LRC) is worst-case-$\mathcal{U}_\delta(\hat{u})$-robust.

Remark 3.9. *For easier distinction of the original program (GCP_u) and the associated local robust counterpart program ($LRC_{\hat{u},\delta}$), we will use u as the general variable in the program (GCP_u), and for the program ($LRC_{\hat{u},\delta}$) we will investigate continuity properties in the variable $\hat{u}$, representing the (moving) center of the local uncertainty set.*

We have already seen in Section 2.3 that the existence of a Slater point is crucial in all the theoretical investigations, hence we need to define the notion of a Slater point for the local robust counterpart program ($LRC_{\hat{u},\delta}$).

Definition 3.10. *A point $x^S \in X$ is a* Slater point *for the local robust counterpart ($LRC_{\hat{u},\delta}$), if*

$$g(x^S,u) <_K 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}).$$

Recall that the notation $g(x^S,u) <_K 0$ means $g(x^S,u) \in \operatorname{int}(-K)$. The subsequent proposition gives an equivalence statement of x^S being a Slater point for the program ($LRC_{\hat{u},\delta}$).

Proposition 3.11. *A point $x^S \in X$ is a Slater point for the local robust counterpart ($LRC_{\hat{u},\delta}$) if and only if there exists an $\varepsilon > 0$ such that*

$$V_\varepsilon\big(g(x^S,u)\big) \subset \operatorname{int}(-K) \quad \forall u \in \mathcal{U}_\delta(\hat{u}).$$

Proof. The backward direction is obvious. The forward direction is proved by contradiction. Assume that for all $\varepsilon > 0$ there exists a $\tilde{u}(\varepsilon) \in \mathcal{U}_\delta(\hat{u})$ such that

$$V_\varepsilon\left(g\big(x^S,\tilde{u}(\varepsilon)\big)\right) \not\subset \operatorname{int}(-K).$$

Thus there exists a point $y(\varepsilon) \in V_\varepsilon(g(x^S, \tilde{u}(\varepsilon)))$ with $y(\varepsilon) \notin \text{int}(-K)$. Consider a sequence $\{\varepsilon_k\}$ with $\varepsilon_k \to 0$ and the associated sequences $\{\tilde{u}_k\} \subset \mathcal{U}_\delta(\hat{u})$ and $\{y_k\}$ with $y_k \in V_{\varepsilon_k}(g(x^S, \tilde{u}_k))$ and $y_k \notin \text{int}(-K)$. Since $\mathcal{U}_\delta(\hat{u})$ is compact, we have without loss of generality $\tilde{u}_k \to \bar{u} \in \mathcal{U}_\delta(\hat{u})$. Furthermore, the sequence $\{y_k\}$ is a Cauchy sequence, since

$$\begin{aligned}\|y_k - y_{k+1}\| \leq \underbrace{\|y_k - g(x^S, \tilde{u}_k)\|}_{<\varepsilon_k} + \underbrace{\|g(x^S, \tilde{u}_k) - g(x^S, \bar{u})\|}_{\to 0} \\ + \underbrace{\|g(x^S, \bar{u}) - g(x^S, \tilde{u}_{k+1})\|}_{\to 0} + \underbrace{\|g(x^S, \tilde{u}_{k+1}) - y_{k+1}\|}_{<\varepsilon_{k+1}}\end{aligned}$$

which tends to 0 for $k \to \infty$. Hence, the sequence $\{y_k\}$ possesses a limit point $\bar{y}$ with $\bar{y} \notin \text{int}(-K)$ as the complement of $\text{int}(-K)$ is a closed set and $y_k \notin \text{int}(-K)$.

Thus, we have

$$\|\bar{y} - g(x^S, \bar{u})\| \leq \underbrace{\|\bar{y} - y_k\|}_{\to 0} + \underbrace{\|y_k - g(x^S, \tilde{u}_k)\|}_{\leq \varepsilon_k} + \underbrace{\|g(x^S, \tilde{u}_k) - g(x^S, \bar{u})\|}_{\to 0}.$$

In the limit $\varepsilon_k \to 0$, the right hand side tends to zero and thus, we eventually get

$$g(x^S, \bar{u}) = \bar{y} \notin \text{int}(-K).$$

Since $\bar{u} \in \mathcal{U}_\delta(\hat{u})$, this is a contradiction to $x^S \in X$ being a Slater point for $(LRC_{\hat{u},\delta})$, i.e. $g(x^S, u) \in \text{int}(-K)$ for all $u \in \mathcal{U}_\delta(\hat{u})$. □

In the next proposition we want to relate the existence of Slater points for the two associated programs – the original program $(GCP_{\hat{u}})$ and the local robust counterpart program $(LRC_{\hat{u},\delta})$.

Proposition 3.12. *Let program $(GCP_{\hat{u}})$ possess a Slater point. Then there exists a $\delta > 0$ such that there exists a Slater point for the local robust counterpart problem $(LRC_{\hat{u},\delta})$.*

Proof. Let $x^S \in X$ denote the Slater point for $(GCP_{\hat{u}})$. Then there exists an $\varepsilon > 0$ such that $V_\varepsilon(g(x^S, \hat{u})) \subset \text{int}(-K)$. Because of g being continuous in u, there exists a $\tilde{\delta} > 0$ such that

$$V_{\frac{\varepsilon}{2}}(g(x^S, u)) \subset \text{int}(-K) \quad \text{for all } u \text{ with } \|u - \hat{u}\| \leq \tilde{\delta},$$

i.e. for all $u \in V_{\tilde{\delta}}(\hat{u})$, a $\tilde{\delta}$-neighborhood around the point $\hat{u}$. Since $\tilde{\delta} > 0$, there exists a δ with $\delta > 0$ such that $\mathcal{U}_\delta(\hat{u}) \subset V_{\tilde{\delta}}(\hat{u})$ and hence x^S is a Slater point for $(LRC_{\hat{u},\delta})$ according to Proposition 3.11. □

Remark 3.13. *If there exist Slater points for (GCP_u) for all $u \in \mathcal{U}$, we also have Slater points for each program $(LRC_{\hat{u},\delta(\hat{u})})$ for sufficiently small $\delta(\hat{u})$ according to Proposition 3.12. As the notation $\delta(\hat{u})$ already indicates, it could be necessary to choose a different size of the local uncertainty set when slighty changing the center point $\hat{u}$. But, using Proposition 3.15 below we will be able to show that there exists one global sizing constant δ_{glob} such that the "same" local uncertainty set can be moved around and we still have a Slater point for each program $(LRC_{\hat{u},\delta_{glob}})$.*

The subsequent result is in analogy to Bliman and Prieur [14].

Proposition 3.14. *Assume that there exists a Slater point for each instance of (GCP_u), i.e. assume*

$$\forall u \in \mathcal{U} \quad \exists\, x^S(u): \quad g\big(x^S(u), u\big) \in \operatorname{int}(-K).$$

Then there exists an $\alpha > 0$ such that for all $u \in \mathcal{U}$ there is an $\bar{x}(u)$ with

$$V_\alpha \left(g\big(\bar{x}(u), u\big)\right) \subset \operatorname{int}(-K).$$

Note that the above expression $V_\alpha \left(g\big(\bar{x}(u), u\big)\right)$ can equivalently be written as

$$V_\alpha \left(g\big(\bar{x}(u), u\big)\right) = g\big(\bar{x}(u), u\big) + \alpha V_1(0)$$

with $V_1(0)$ denoting a 1-neighborhood around the origin, i.e. some "normed neighborhood", similar to the unit ball.

Proof. Define

$$Z_k := \{u \in \mathcal{U} \mid \exists x \in X : V_{\frac{1}{k}}\big(g(x,u)\big) \subset \operatorname{int}(-K)\}.$$

Thus, with this definition we have to show that there exists $\hat{k}$ such that $Z_k = \mathcal{U}$ for all $k \geq \hat{k}$. With $\frac{1}{\infty} := 0$ it is obvious that the limit $Z_\infty = \mathcal{U}$ since we assumed the existence of a Slater point for each instance of (GCP_u). Furthermore, for k sufficiently large we also have that $Z_k \neq \emptyset$.

We will proof the existence of a $\hat{k}$ such that $Z_k = \mathcal{U}$ for all $k \geq \hat{k}$ by contradiction. Assume that for all k, the set Z_k is not equal to $\mathcal{U}$. Then there exists a sequence $\{u_k\} \in \mathcal{U} \setminus Z_k$. Since $\mathcal{U}$ is compact, this sequence has an accumulation point in $\mathcal{U}$, say $\bar{u}$, and without loss of generality $\bar{u} = \lim\limits_{k\to\infty} u_k$. By assumption, problem $(GCP_{\bar{u}})$ has a Slater point, i.e. there exists $x^S(\bar{u})$ such that

$$g(x^S(\bar{u}), \bar{u}) \in \operatorname{int}(-K).$$

Thus, there exists an $\varepsilon > 0$ such that

$$V_\varepsilon \left(g\big(x^S(\bar{u}), \bar{u}\big)\right) \subset \operatorname{int}(-K).$$

Since g is continuous in $\bar{u}$, it holds that for each $\beta > 0$ there exists a $\gamma > 0$ with

$$\|g(x^S(\bar{u}), \bar{u}) - g(x^S(\bar{u}), u)\| < \beta \quad \text{for all } u \text{ with } \|\bar{u} - u\| < \gamma.$$

Thus, especially for $\beta = \dfrac{\varepsilon}{2}$ there exists a $\gamma > 0$ with

$$\|g(x^S(\bar{u}), \bar{u}) - g(x^S(\bar{u}), u)\| < \frac{\varepsilon}{2} \quad \text{for all } u \text{ with } \|\bar{u} - u\| < \gamma.$$

For sufficiently large k, we have $\|\bar{u} - u_k\| < \gamma$, thus

$$\|g(x^S(\bar{u}), \bar{u}) - g(x^S(\bar{u}), u_k)\| < \frac{\varepsilon}{2}$$

which then implies

$$V_{\frac{\varepsilon}{2}}\left(g\big(x^S(\bar{u}), u_k\big)\right) \subset \operatorname{int}(-K) \quad \text{for all } u_k \text{ with } k \text{ large.}$$

Choosing $\bar{k}$ such that both $\|u_{\bar{k}} - \bar{u}\| < \gamma$ and $\dfrac{1}{\bar{k}} < \dfrac{\varepsilon}{2}$, we can conclude that

$$V_{\frac{1}{\bar{k}}}\left(g\big(x^S(\bar{u}), u_{\bar{k}}\big)\right) \subset \operatorname{int}(-K).$$

But this in turn implies that $u_{\bar{k}} \in Z_{\bar{k}}$ which contradicts the assumption. Thus, there exists $\hat{k}$ such that $Z_k = \mathcal{U}$ for all $k \geq \hat{k}$, and hence we have the existence of an $\alpha > 0$ such that for all $u \in \mathcal{U}$ there exists an $\bar{x}(u)$ with

$$V_\alpha\left(g\big(\bar{x}(u), u\big)\right) \subset \operatorname{int}(-K). \qquad \square$$

It is worth stressing that the sizing constant α in Proposition 3.14 does not depend on the parameter u which enables us to prove the next statement.

Proposition 3.15. *Assume the existence of a Slater point for (GCP_u) for all $u \in \mathcal{U}$, and let g be globally Lipschitz continuous in u. Then there exists a global size $\delta_{glob} > 0$ such that the local robust counterpart program ($LRC_{\hat{u},\delta_{glob}}$) for any $\hat{u} \in \operatorname{int}(\mathcal{U})$ possesses again a Slater point.*

Proof. Having the existence of a Slater point for (GCP_u) for all $u \in \mathcal{U}$, Proposition 3.14 gives the existence of an $\alpha > 0$ such that for all $\hat{u} \in \operatorname{int}(\mathcal{U})$ there exists $x^S(\hat{u})$ with $V_\alpha(g(x^S(\hat{u}), \hat{u})) \subset \operatorname{int}(-K)$. Using global Lipschitz continuity of g at $\hat{u}$, we have for all $\hat{u} \in \operatorname{int}(\mathcal{U})$

$$V_{\frac{\alpha}{2}}(g(x^S(\hat{u}), u)) \subset \operatorname{int}(-K) \quad \forall u \in V_{\frac{\alpha/2}{L}}(\hat{u})$$

with $L > 0$ being the global Lipschitz constant of g.
Finally, defining $\delta_{glob} > 0$ such that $\mathcal{U}_{\delta_{glob}}(\hat{u}) \subset V_{\frac{\alpha/2}{L}}(\hat{u})$, the proof is complete. $\square$

3.2 Stability of the LRC

After introducing the local robust counterpart and providing the necessary notations, we now investigate the program $(LRC_{\hat{u},\delta})$ with respect to its continuity characteristics – analogous to Section 2.3 where we have dealt with the program (GCP_u).

In contrast to the original problem (GCP_u) we now have two parameters[3] defining the local robust counterpart program $(LRC_{\hat{u},\delta})$: the center of the local uncertainty set, $\hat{u}$, and its size δ. Thus, we will denote the associated mappings of the program $(LRC_{\hat{u},\delta})$ as follows:

Notation 3.16.

- *The feasible set mapping* $\mathcal{F}_{LRC} : \mathcal{U} \times \mathbb{R}_+ \to \mathcal{P}(X)$ *is determined by the according feasibility sets for each parameter pair* $(\hat{u}, \delta)$*,*

$$\mathcal{F}_{LRC}(\hat{u}, \delta) := \{x \in X \mid g(x, u) \leq_K 0 \ \ \forall u \in \mathcal{U}_\delta(\hat{u})\}.$$

- *The set of Slater points of the program* $(LRC_{\hat{u},\delta})$ *is given by*

$$\mathcal{F}^S_{LRC}(\hat{u}, \delta) := \{x \in X \mid g(x, u) <_K 0 \ \ \forall u \in \mathcal{U}_\delta(\hat{u})\}.$$

- *The objective function of the local robust counterpart program* $(LRC_{\hat{u},\delta})$ *is denoted by* $f_{LRC} : \mathbb{R}^n \times \mathcal{U} \times \mathbb{R}_+ \to \mathbb{R}$ *and is defined by*

$$f_{LRC}(x, \hat{u}, \delta) := \max_{u \in \mathcal{U}_\delta(\hat{u})} f(x, u).$$

- *The extreme value function or optimal value function* $f^*_{LRC} : \mathcal{U} \times \mathbb{R}_+ \to \mathbb{R}$ *is*

$$f^*_{LRC}(\hat{u}, \delta) := \min\{f_{LRC}(x, \hat{u}, \delta) \mid x \in \mathcal{F}_{LRC}(\hat{u}, \delta)\}.$$

- *The optimal set mapping* $\mathcal{F}^*_{LRC} : \mathcal{U} \times \mathbb{R}_+ \to \mathcal{P}(X)$ *is given by the sets of optimal solutions,*

$$\mathcal{F}^*_{LRC}(\hat{u}, \delta) := \{x \in \mathcal{F}_{LRC}(\hat{u}, \delta) \mid f_{LRC}(x, \hat{u}, \delta) \leq f^*_{LRC}(\hat{u}, \delta)\}.$$

To be able to investigate the same stability characterizations as in Section 2.3, we need to verify the necessary prerequisites of Assumption 2.14 to apply the already proved results. Hence, it remains to show continuity and convexity of the robust objective function and the new (semi-infinite) constraint.

[3] In the cases where we will deal additionally with different *shapes* of the uncertainty set, we will explicitly point this out at the appropriate places.

Lemma 3.17.

(i) The robustified objective function $f_{LRC}(x, \hat{u}, \delta) = \max_{u \in \mathcal{U}_\delta(\hat{u})} f(x, u)$ *is continuous in* $(x, \hat{u}, \delta)$.

(ii) Furthermore, $f_{LRC}(x, \hat{u}, \delta)$ *is convex in* x *for fixed* $(\hat{u}, \delta)$*, convex in* $\hat{u}$ *for fixed* (x, δ) *and convex and monotonically increasing in* δ *for fixed* $(x, \hat{u})$.

(iii) If $f(x, u)$ *is strictly convex in* x *for all* $u \in \mathcal{U}$*, then* $f_{LRC}(x, \hat{u}, \delta)$ *is again strictly convex in* x.

Proof.

(i) To show continuity of f_{LRC}, define $z := (x, \hat{u}, \delta)$, let $v \in \mathcal{U}$ and consider the following parametric program:

$$\max_{v \in \mathcal{F}_P(z)} h(v, z) \tag{P_z}$$

with $h(v, z) = f(x, v)$ and $\mathcal{F}_P(z) = \mathcal{U}_\delta(\hat{u})$. Note that

$$f_{LRC}(z) = f_{LRC}(x, \hat{u}, \delta) = \max_{u \in \mathcal{U}_\delta(\hat{u})} f(x, u) = \max_{v \in \mathcal{F}_P(z)} h(v, z) = f_P^*(z)$$

is the optimal value function f_P^* of the auxiliary problem (P_z). As we now want to apply the already established result about continuity of the extreme value function, Proposition 2.40, we note the following:

- The function h is continuous in z and v by definition. (Note that convexity of the objective function is not needed to prove continuity of the optimal value function.)
- The feasibility set of (P_z), $\mathcal{F}_P(z) = \mathcal{U}_\delta(\hat{u})$ is compact by definition of the local uncertainty set.
- Hausdorff continuity of the mapping $\mathcal{F}_P : \mathbb{R}^n \times \mathcal{U} \times \mathbb{R}_+ \to \mathbb{R}^d$ follows from the definition of $\mathcal{F}_P(z) = \mathcal{U}_\delta(\hat{u}) = \hat{u} + \delta\mathcal{U}'$.

Thus, Proposition 2.40 is applicable and yields continuity of f_{LRC} in $z = (x, \hat{u}, \delta)$.

(ii) For fixed $(\hat{u}, \delta)$, the feasibility set $\mathcal{F}_P(z)$ is constant and thus f_{LRC} is convex in x as the pointwise maximum of convex functions, see e.g. Rockafellar [70], Theorem 5.5.

For fixed (x, δ), convexity of f_{LRC} in $\hat{u}$ is shown using its definition and convexity of f in u. Note that we can rewrite f_{LRC} as

$$\begin{aligned} f_{LRC}(x, \hat{u}, \delta) &= \max_{u \in \mathcal{U}_\delta(\hat{u})} f(x, u) \\ &= \max_{w \in \mathcal{U}_\delta(0)} f(x, \hat{u} + w). \end{aligned}$$

Thus we get

$$\begin{aligned}
f_{LRC}\big(x, \alpha\hat{u}_1 + (1-\alpha)\hat{u}_2, \delta\big) = \\
&= \max_{w \in \mathcal{U}_\delta(0)} f(x, \alpha\hat{u}_1 + (1-\alpha)\hat{u}_2 + w) \\
&= \max_{\substack{w_1, w_2 \in \mathcal{U}_\delta(0) \\ w_1 = w_2}} f\big(x, \alpha\hat{u}_1 + (1-\alpha)\hat{u}_2 + \alpha w_1 + (1-\alpha)w_2\big) \\
&\leq \max_{\substack{w_1, w_2 \in \mathcal{U}_\delta(0) \\ w_1 = w_2}} \alpha f(x, \hat{u}_1 + w_1) + (1-\alpha) f(x, \hat{u}_2 + w_2) \\
&\leq \max_{w_1, w_2 \in \mathcal{U}_\delta(0)} \alpha f(x, \hat{u}_1 + w_1) + (1-\alpha) f(x, \hat{u}_2 + w_2) \\
&\leq \max_{w_1 \in \mathcal{U}_\delta(0)} \alpha f(x, \hat{u}_1 + w_1) + \max_{w_2 \in \mathcal{U}_\delta(0)} \alpha f(x, \hat{u}_2 + w_2) \\
&= \alpha f_{LRC}(x, \hat{u}_1, \delta) + (1-\alpha) f_{LRC}(x, \hat{u}_2, \delta).
\end{aligned}$$

Monotonicity of $f_{LRC}(x, \hat{u}, \delta)$ in δ for fixed $(x, \hat{u})$ follows straightforwardly from the definition since for a shrinking feasibility set the maximum value can at most be equal or is decreasing otherwise. To prove convexity in δ, we first note that for all $0 \leq \alpha \leq 1$ we can represent each $w \in \mathcal{U}_{\alpha\delta_1 + (1-\alpha)\delta_2}(0)$ as

$$w = \alpha w_1 + (1-\alpha) w_2 \quad \text{with } w_1 \in \mathcal{U}_{\delta_1}(0), w_2 \in \mathcal{U}_{\delta_2}(0).$$

Thus,

$$\begin{aligned}
f_{LRC}\big(x, \hat{u}, \alpha\delta_1 + (1-\alpha)\delta_2\big) = \\
&= \max_{u \in \mathcal{U}_{\alpha\delta_1 + (1-\alpha)\delta_2}(\hat{u})} f(x, u) \\
&= \max_{w \in \mathcal{U}_{\alpha\delta_1 + (1-\alpha)\delta_2}(0)} f(x, \hat{u} + w) \\
&= \max_{\substack{w_1 \in \mathcal{U}_{\delta_1}(0) \\ w_2 \in \mathcal{U}_{\delta_2}(0)}} f\big(x, \alpha\hat{u} + (1-\alpha)\hat{u} + \alpha w_1 + (1-\alpha)w_2\big) \\
&\leq \max_{\substack{w_1 \in \mathcal{U}_{\delta_1}(0) \\ w_2 \in \mathcal{U}_{\delta_2}(0)}} \alpha f(x, \hat{u} + w_1) + (1-\alpha) f(x, \hat{u} + w_2) \\
&\leq \max_{w_1 \in \mathcal{U}_{\delta_1}(0)} \alpha f(x, \hat{u} + w_1) + \max_{w_2 \in \mathcal{U}_{\delta_2}(0)} (1-\alpha) f(x, \hat{u} + w_2) \\
&= \alpha f_{LRC}(x, \hat{u}, \delta_1) + (1-\alpha) f_{LRC}(x, \hat{u}, \delta_2).
\end{aligned}$$

(iii) Strict convexity of $f(x, u)$ in x gives

$$f\big(\alpha x + (1-\alpha)y, u\big) < \alpha f(x, u) + (1-\alpha) f(y, u)$$

for $x, y \in X, x \neq y$ and $0 < \alpha < 1$ and fixed u. Hence, it holds as well that

$$f\big(\alpha x + (1-\alpha)y, u\big) < \max_{v \in \mathcal{U}_\delta(\hat{u})} \alpha f(x, v) + (1-\alpha) f(y, v)$$

and finally, as this inequality is valid for all $u \in \mathcal{U}_\delta(\hat{u})$, we have

$$\begin{aligned}\max_{u\in\mathcal{U}_\delta(\hat{u})} f\big(\alpha x + (1-\alpha)y, u\big) &< \max_{v\in\mathcal{U}_\delta(\hat{u})} \alpha f(x,v) + (1-\alpha) f(y,v) \\ &\leq \max_{v\in\mathcal{U}_\delta(\hat{u})} \alpha f(x,v) + \max_{v\in\mathcal{U}_\delta(\hat{u})} (1-\alpha) f(y,v).\end{aligned}$$

Thus,

$$f_{LRC}\big(\alpha x + (1-\alpha)y, \hat{u}, \delta\big) < \alpha f_{LRC}(x,\hat{u},\delta) + (1-\alpha) f_{LRC}(y,\hat{u},\delta) \qquad \square$$

Thus, so far we have established continuity and convexity of the objective function. Next we investigate properties of the robust constraint. As the formulation "for all $u \in \mathcal{U}_\delta(\hat{u})$" is difficult to handle when it comes to continuity and convexity, we first reformulate the original semi-infinite constraint to a single real-valued constraint. This will greatly simplify the proofs of the needed properties.

Lemma 3.18.

(i) Let

$$\mathcal{F}^1_{LRC}(\hat{u},\delta) = \{x \in X \mid g(x,u) \leq_K 0 \ \ \forall u \in \mathcal{U}_\delta(\hat{u})\}$$

and

$$\mathcal{F}^2_{LRC}(\hat{u},\delta) = \{x \in X \mid G(x,\hat{u},\delta) \leq 0\}$$

with

$$G(x,\hat{u},\delta) := \max_{\substack{u\in\mathcal{U}_\delta(\hat{u})\\ \lambda\in K^*\\ \|\lambda\|=1}} \lambda^T g(x,u).$$

It then holds that $\mathcal{F}^1_{LRC}(\hat{u},\delta) = \mathcal{F}^2_{LRC}(\hat{u},\delta)$.

(ii) A point $x^S \in X$ *is a Slater point for the program* $(LRC_{\hat{u},\delta})$ *if and only if*

$$G(x^S,\hat{u},\delta) < 0,$$

i.e.

$$\begin{aligned}\mathcal{F}^S_{LRC}(\hat{u},\delta) &= \{x \in X \mid g(x,u) <_K 0 \ \ \forall u \in \mathcal{U}_\delta(\hat{u})\} \\ &= \{x \in X \mid G(x,\hat{u},\delta) < 0\}.\end{aligned}$$

Proof.

(i) To prove the forward direction, let $\hat{x} \in \mathcal{F}^1_{LRC}(\hat{u}, \delta)$, i.e. it holds that

$$g(\hat{x}, u) \leq_K 0 \quad \text{for all } u \in \mathcal{U}_\delta(\hat{u}).$$

Using Lemma 2.10 gives

$$\lambda^T g(\hat{x}, u) \leq 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}), \forall \lambda \in K^*.$$

Considering only a subset of K^*, the statement of course remains valid, thus

$$\lambda^T g(\hat{x}, u) \leq 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}), \forall \lambda \in K^* \text{ with } \|\lambda\| = 1$$

which implies

$$\max_{\substack{u \in \mathcal{U}_\delta(\hat{u}) \\ \lambda \in K^* \\ \|\lambda\|=1}} \lambda^T g(\hat{x}, u) \leq 0.$$

As this program is the definition of $G(x, \hat{u}, \delta)$ from above, we have

$$G(\hat{x}, \hat{u}, \delta) \leq 0,$$

thus, $\hat{x} \in \mathcal{F}^2_{LRC}(\hat{u}, \delta)$.

To prove the backward direction, let $\hat{x} \in \mathcal{F}^2_{LRC}(\hat{u}, \delta)$. Thus, it holds that

$$\max_{\substack{u \in \mathcal{U}_\delta(\hat{u}) \\ \lambda \in K^* \\ \|\lambda\|=1}} \lambda^T g(\hat{x}, u) \leq 0$$

which implies

$$\lambda^T g(\hat{x}, u) \leq 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}), \forall \lambda \in K^* \text{ with } \|\lambda\| = 1.$$

Incorporating the condition $\|\lambda\| = 1$ into the inequality yields

$$\frac{\lambda^T}{\|\lambda\|} g(\hat{x}, u) \leq 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}), \forall \lambda \in K^*,$$

and as the inequality furthermore remains unaffected by multiplication with a strictly positive number, we obtain

$$\|\lambda\| \frac{\lambda^T}{\|\lambda\|} g(\hat{x}, u) = \lambda^T g(\hat{x}, u) \leq 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}), \forall \lambda \in K^*.$$

Lemma 2.10 thus gives that

$$g(\hat{x}, u) \leq_K 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}),$$

i.e. $\hat{x} \in \mathcal{F}^1_{LRC}(\hat{u}, \delta)$.

(ii) For the forward direction, let $x^S \in X$ be a Slater point for $(LRC_{\hat{u},\delta})$, i.e. it holds that

$$g(x^S, u) <_K 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}).$$

From Lemma 2.11 we get the strict inequality

$$\lambda^T g(x^S, u) < 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}), \forall \lambda \in K^* \backslash \{0\}.$$

We again reduce the scope of the statement to a subset of K^* and obtain

$$\lambda^T g(x^S, u) < 0 \;\; \forall u \in \mathcal{U}_\delta(\hat{u}), \forall \lambda \in K^*, \|\lambda\| = 1.$$

Since the set

$$\{(\lambda, u) \mid \lambda \in K^*, \|\lambda\| = 1, u \in \mathcal{U}_\delta(\hat{u})\} = \{(\lambda, u) \mid \lambda \in K^*, \|\lambda\| = 1\} \times \mathcal{U}_\delta(\hat{u})$$

is compact ($\|\lambda\| = 1$ yields compactness of λ and $\mathcal{U}_\delta(\hat{u})$ is a compact set), the maximum of the continuous function $\lambda^T g(x^S, u)$ is attained, i.e. the above statement is equivalent to

$$\max_{\substack{u \in \mathcal{U}_\delta(\hat{u}) \\ \lambda \in K^* \\ \|\lambda\| = 1}} \lambda^T g(x^S, u) < 0,$$

thus $G(x_s, \hat{u}, \delta) < 0$.

To prove the backward direction we proceed analogously to part (i), using the strict inequality and excluding $\lambda = 0$. Thus we reach the point where it holds

$$\lambda^T g(x^S, u) < 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}), \forall \lambda \in K^* \backslash \{0\}$$

from where we straightforwardly get

$$g(x^S, u) <_K 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}).$$

applying Lemma 2.11. □

Remark 3.19. *Note that the statement in part (i) of Lemma 3.18 would remain unchanged even if the feasibility set of the auxiliary optimization problem to define the function G was relaxed to $\{(\lambda, u) \mid u \in \mathcal{U}_\delta(\hat{u}), \lambda \in K^*\}$. The restriction $\|\lambda\| = 1$ was added for two reasons: one is to achieve compactness of the feasibility set, which is needed both in part (ii) of the lemma and in the subsequent proof of Lemma 3.20; the other one is to exclude the case $\lambda = 0$ which is necessary to deal with the equivalence statement of a Slater point, part (ii) of the above lemma.*

The equivalence statement in Lemma 3.18, part (i) thus allows the reformulation of the semi-infinite constraint in the local robust counterpart program into a single real-valued constraint by interpretation as the optimal value function of an optimization problem. Hence, the local robust counterpart is given by

$$\begin{aligned} \min_{x \in X} \quad & \max_{u \in \mathcal{U}_\delta(\hat{u})} f(x,u) \\ & G(x,\hat{u},\delta) \leq 0 \end{aligned} \tag{$LRC_{\hat{u},\delta}$}$$

with $G(x,\hat{u},\delta)$ defined by

$$G(x,\hat{u},\delta) := \max_{\substack{u \in \mathcal{U}_\delta(\hat{u}) \\ \lambda \in K^* \\ \|\lambda\|=1}} \lambda^T g(x,u).$$

As we have already established continuity and convexity of the objective function $f_{LRC}(\hat{u},\delta)$ in Proposition 3.17, it remains to show continuity and convexity of the new constraint $G(x,\hat{u},\delta)$ before being able to apply the results from Section 2.3 to the robust program.

Lemma 3.20. *The function $G : \mathbb{R}^n \times \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}$ as defined in Lemma 3.18 is*

(i) (jointly) continuous in $(x,\hat{u},\delta)$,

(ii) convex in x for fixed $(\hat{u},\delta)$, convex in $\hat{u}$ for fixed (x,δ) and convex and monotonically increasing in δ for fixed $(x,\hat{u})$.

Proof. Consider the auxiliary optimization problem

$$\min_{\substack{u \in \mathcal{U}_\delta(\hat{u}) \\ \lambda \in K^* \\ \|\lambda\|=1}} -\lambda^T g(x,u) \tag{P_{aux}}$$

where λ and u are the variables and $x, \hat{u}$ and δ represent the parameters.

(i) We want to prove continuity of the optimal value function $f^*_{aux} = G$ with respect to the parameters using Proposition 2.40. Let $z := (x,\hat{u},\delta) \in Z := X \times \mathcal{U} \times \mathbb{R}_+$ denote an arbitrary choice of parameters. Thus, we need to verify the prerequisites that the feasibility set $\mathcal{F}_{aux}$ is H-continuous at z and the objective function f_{aux} is continuous on $\mathcal{F}_{aux}(r) \times Z$. Note that convexity of f_{aux} and $\mathcal{F}_{aux}$ is not required in this case.

- The feasibility set $\mathcal{F}_{aux}(z) \subset \mathbb{R}^m \times \mathbb{R}^d$ for the chosen point z is given by

$$\mathcal{F}_{aux}(z) = \mathcal{F}_{aux}(x,\hat{u},\delta) = \{\lambda \in K^* \mid \|\lambda\| = 1\} \times \mathcal{U}_\delta(\hat{u}).$$

  Hence, the mapping $\mathcal{F}_{aux}$ is H-continuous at z due to the definition of $\mathcal{U}_\delta(\hat{u}) = \hat{u} + \delta\mathcal{U}'$. Note also that $\mathcal{F}_{aux}(z)$ is a compact set for each z.

- The objective function of the auxiliary problem with the variables (λ, u) and for the parameter $z = (x, \hat{u}, \delta)$ is given by

$$f_{aux}(\lambda, u, x, \hat{u}, \delta) = -\lambda^T g(x, u)$$

which is continuous for all $(\lambda, u) \in \mathcal{F}_{aux}(x, \hat{u}, \delta)$.

Having these prerequisites we can use Proposition 2.40 to conclude continuity of $f^*_{aux} = G$ in $z = (x, \hat{u}, \delta)$.

(ii) To show convexity of $G(., \hat{u}, \delta)$, we first note that since $g(., u)$ is K-convex, $\lambda^T g(., u)$ is convex for any $\lambda \in K^*$, see Proposition 2.13. Then, $G(., \hat{u}, \delta)$ is convex as the pointwise maximum of convex functions, see e.g. Rockafellar [70], Theorem 5.5.

Convexity of $G(x, ., \delta)$ follows using the same arguments.

Monotonicity of $G(x, \hat{u}, \delta)$ in δ for fixed $(x, \hat{u})$ follows straightforwardly from the definition, using the same argument as in Lemma 3.17. Convexity of $G(x, \hat{u}, \delta)$ in δ is proved analogously to convexity of $f_{LRC}(x, \hat{u}, \delta)$ in δ:

$$\begin{aligned}
&G(x, \hat{u}, \alpha\delta_1 + (1-\alpha)\delta_2) = \\
&\qquad = \max_{\substack{u \in \mathcal{U}_{\alpha\delta_1 + (1-\alpha)\delta_2}(\hat{u}) \\ \lambda \in K^* \\ \|\lambda\| = 1}} \lambda^T g(x, u) \\
&\qquad = \max_{\substack{w \in \mathcal{U}_{\alpha\delta_1 + (1-\alpha)\delta_2}(0) \\ \lambda \in K^* \\ \|\lambda\| = 1}} \lambda^T g(x, \hat{u} + w) \\
&\qquad = \max_{\substack{w_1 \in \mathcal{U}_{\delta_1}(0) \\ w_2 \in \mathcal{U}_{\delta_2}(0) \\ \lambda \in K^* \\ \|\lambda\| = 1}} \lambda^T g\big(x, \alpha\hat{u} + (1-\alpha)\hat{u} + \alpha w_1 + (1-\alpha) w_2\big) \\
&\qquad \leq \max_{\substack{w_1 \in \mathcal{U}_{\delta_1}(0) \\ w_2 \in \mathcal{U}_{\delta_2}(0) \\ \lambda \in K^* \\ \|\lambda\| = 1}} \alpha\lambda^T g(x, \hat{u} + w_1) + (1-\alpha)\lambda^T g(x, \hat{u} + w_2) \\
&\qquad \leq \max_{\substack{w_1 \in \mathcal{U}_{\delta_1}(0) \\ \lambda \in K^* \\ \|\lambda\| = 1}} \alpha\lambda^T g(x, \hat{u} + w_1) + \max_{\substack{w_2 \in \mathcal{U}_{\delta_2}(0) \\ \lambda \in K^* \\ \|\lambda\| = 1}} (1-\alpha)\lambda^T g(x, \hat{u} + w_2) \\
&\qquad = \alpha G(x, \hat{u}, \delta_1) + (1-\alpha) G(x, \hat{u}, \delta_2). \qquad \square
\end{aligned}$$

The results of Lemma 3.17 and Lemma 3.20 showed that the general requirements for the convex conic optimization problem (GCP_u) summarized in Assumption 2.14 also hold for the local robust counterpart problem $(LRC_{\hat{u},\delta})$ Thus, now that we especially have both continuity and convexity of the robust objective and constraint, we can prove the same stability properties for the local

robust counterpart program as we did for the original convex problem applying the already established results from Section 2.3.

Theorem 3.21. *Consider the local robust counterpart problem ($LRC_{\hat{u},\delta}$) with the associated feasible set mapping $\mathcal{F}_{LRC}$, the extreme value function f^*_{LRC} and the optimal set mapping $\mathcal{F}^*_{LRC}$. The following statements hold:*

(i) The mapping $\mathcal{F}_{LRC}$ is closed and Hausdorff upper semicontinuous for all $(\hat{u}, \delta) \in \mathcal{U} \times \mathbb{R}_+$.

(ii) Let ($LRC_{\bar{u},\bar{\delta}}$) possess a Slater point. Then the feasible set mapping $\mathcal{F}_{LRC}$ is Hausdorff continuous at $(\bar{u}, \bar{\delta})$.

(iii) Let $\mathcal{F}_{LRC}$ be Hausdorff continuous at $(\bar{u}, \bar{\delta})$. Then the optimal value function f^*_{LRC} is continuous at $(\bar{u}, \bar{\delta})$.

(iv) Let $\mathcal{F}_{LRC}$ be Hausdorff continuous at $(\bar{u}, \bar{\delta})$. Then the optimal set mapping $\mathcal{F}^*_{LRC}$ is closed at $(\bar{u}, \bar{\delta})$ and Hausdorff upper semicontinuous at $(\bar{u}, \bar{\delta})$.

(v) Let $\mathcal{F}_{LRC}$ be Hausdorff continuous at $(\bar{u}, \bar{\delta})$ and let $\mathcal{F}^*_{LRC}(\bar{u}, \bar{\delta})$ be a singleton. Then the optimal set mapping $\mathcal{F}^*_{LRC}$ is Hausdorff continuous at $(\bar{u}, \bar{\delta})$.

Proof.

(i) The mapping $\mathcal{F}_{LRC}$ is closed since the function G is continuous (see Lemma 3.20). Furthermore, closedness of $\mathcal{F}_{LRC}$ together with X being compact yields Hausdorff upper semicontinuity according to Lemma 2.26 (i).

(ii) Having closedness of $\mathcal{F}_{LRC}$ at $(\bar{u}, \bar{\delta})$ and continuity and convexity of f_{LRC} and G (see Lemmas 3.17 and 3.20), the statement follows directly from Proposition 2.38 together with the existence of a Slater point.

(iii) Continuity of both f_{LRC} and $\mathcal{F}_{LRC}$ at $(\bar{u}, \bar{\delta})$ and compactness of $\mathcal{F}_{LRC}(\bar{u}, \bar{\delta})$ imply continuity of $f^*_{LRC}(\hat{u}, \delta)$ according to Proposition 2.40.

(iv) Closedness of $\mathcal{F}^*_{LRC}$ follows from Theorem 2.29 together with part (iii). Hausdorff upper semicontinuity follows directly from Proposition 2.50.

(v) Hausdorff continuity in the case of a unique solution is given according to Theorem 2.45. □

Theorem 3.21 shows that the program ($LRC_{\hat{u},\delta}$) itself possesses analogous continuity characteristics as the original program (GCP_u) with respect to the uncertainty parameters $\hat{u}$ and δ. Hence, when robustifying the original problem to the local robust counterpart, we do not lose any stability properties. This means especially that the existence of Slater point for ($LRC_{\hat{u},\delta}$) – which is closely linked to the existence of a Slater point for (GCP_u) – suffices to assure Hausdorff

continuity of the feasible set mapping which then also implies continuity of the extreme value function and at least Hausdorff upper semicontinuity of the optimal set mapping.

In the following, we are especially interested in the connection of the original problem $(GCP_{\hat{u}})$ to the robust program $(LRC_{\hat{u},\delta})$, expressed in the limit point $\bar{\delta} = 0$, since this reduces the local robust counterpart program to the original problem $(GCP_{\hat{u}})$. We expect that for $\delta \to 0$ the sequence of robust solutions converges to an optimal solution of the original program. Or, more precisely – since we do not have uniqueness of the optimal solution – we expect that the sequence of sets of optimal solutions to the robust problem converges to a subset of the set of optimal solutions of the original problem. This result is stated in the following corollary.

Corollary 3.22. *Let $\bar{u} \in \mathcal{U}$ be fixed and assume the existence of a Slater point for the program $(GCP_{\bar{u}})$. Let $x^*_{LRC}(\bar{u}, \delta)$ denote an optimal solution to the corresponding local robust counterpart program $(LRC_{\bar{u},\delta})$. Then it holds:*

(i) The optimal set mapping $\mathcal{F}^*_{LRC}$ is Hausdorff upper semicontinuous at $\bar{\delta} = 0$, i.e. every accumulation point of a sequence $\{x^*_{LRC}(\bar{u}, \delta_k)\}$ with

$$x^*_{LRC}(\bar{u}, \delta_k) \in \mathcal{F}^*_{LRC}(\bar{u}, \delta_k)$$

and $\delta_k \to 0$ is in $\mathcal{F}^*_{LRC}(\bar{u}, 0) = \mathcal{F}^*(\bar{u})$.

(ii) If furthermore $\mathcal{F}^*_{LRC}(\bar{u}, 0)$ is a singleton, the mapping $\mathcal{F}^*_{LRC}$ is Hausdorff continuous at $\bar{\delta} = 0$, i.e. the limit of the sequence $\{x^*_{LRC}(\bar{u}, \delta_k)\}$ exists and the limit point is an optimal solution to $(GCP_{\bar{u}})$.

Proof. According to Proposition 3.12 the existence of a Slater point for $(GCP_{\bar{u}})$ implies the existence of a Slater point for $(LRC_{\bar{u},\delta_k})$ with δ_k sufficiently close to $\bar{\delta} = 0$. Thus, for small enough δ_k and especially for $\bar{\delta} = 0$ we have the necessary prerequisites for $f_{LRC}, \mathcal{F}_{LRC}$ and f^*_{LRC} being (Hausdorff) continuous at $\bar{\delta} = 0$. Then, (i) and (ii) follow directly from part (iv) and (v) of Theorem 3.21. □

In the following examples we want to illustrate the results of Corollary 3.22 and also its limitations. The first example shows Hausdorff upper semicontinuity, i.e. that the sequence of robust optimal solutions tends to an optimal solution of the original problem. In that particular example the feasibility sets of both problems coincide and are the constant interval $[0, 1]$, hence the feasible set mappings are Hausdorff continuous. Since the existence of a Slater point of the original program is only necessary to assure Hausdorff continuity of $\mathcal{F}^*$ and $\mathcal{F}^*_{LRC}$, this requirement can be dropped in cases where $\mathcal{F}^* = \mathcal{F}^*_{LRC} =$ constant.

Example 3.23. *Consider the optimization problem*

$$\min_{x\in[0,1]} \quad ux \tag{P}$$

with $\mathcal{U} = [-1, 1]$ *and the particular parameter* $\bar{u} = 0$*. The local uncertainty set is hence* $\mathcal{U}_\delta(0) = [-\delta, \delta]$*. Similar to Example 2.43 the optimal set mapping of problem (P) is given by*

$$\mathcal{F}^*(u) = \begin{cases} \{1\} & u < 0 \\ [0,1] & u = 0 \\ \{0\} & u > 0. \end{cases}$$

The corresponding local robust counterpart to (P) formulates to

$$\min_{x\in[0,1]} \max_{u\in\mathcal{U}_\delta(0)} \quad ux = \min_{x\in[0,1]} \quad \delta x$$

since $\max_{u\in\mathcal{U}_\delta(0)} \quad ux = \delta x$ *for* $x \geq 0$*. Hence, the optimal set mapping of the robust problem at the point* $\bar{u} = 0$ *is given by*

$$\mathcal{F}^*_{LRC}(0, \delta) = \begin{cases} [0,1] & \delta = 0 \\ \{0\} & \delta > 0. \end{cases}$$

Thus, it holds that

$$\mathcal{F}^*_{LRC}(0, \delta) \ni x^*_{LRC}(0, \delta) \quad \rightarrow \quad x^*(0) \in \mathcal{F}^*_{LRC}(0, 0) = \mathcal{F}^*(0),$$

i.e. $\mathcal{F}^*_{LRC}$ *is Hausdorff upper semicontinuous at* $(0, 0)$*.*

This first example hence shows that Hausdorff continuity of the feasible set mappings guarantees Hausdorff upper semicontinuity of the optimal set mapping in $\delta = 0$. We do not get Hausdorff lower semicontinuity, since the set of optimal solutions of the original problem at $\bar{u} = 0$ is not a singleton. The second example illustrates that Hausdorff upper semicontinuity of $\mathcal{F}_{LRC}$ (in δ) does not suffice to assure Hausdorff upper semicontinuity of $\mathcal{F}^*_{LRC}$. Hence, the prerequisite of having a Slater point is relevant for the results stated in Corollary 3.22.

Example 3.24. *Consider the optimization problem*

$$\begin{aligned} \min_{x\in[-1,1]} \quad & (x-1)^2 \\ s.t. \quad & ux \leq 0 \end{aligned} \tag{P}$$

with $u \in \mathcal{U} = [-1, 1]$*. Let the local uncertainty set be given by* $\mathcal{U}_\delta(\hat{u}) = [\hat{u}-\delta, \hat{u}+\delta]$ *with* $\delta > 0$ *and such that* $(\hat{u}, \delta)$ *is admissible. As the objective function is independent of the uncertain parameter* u*, robustification only affects the constraint,*

and the local robust counterpart to program (P) is thus described by

$$\begin{aligned} \min_{x\in[-1,1]} \quad & (x-1)^2 \\ s.t. \quad & \hat{u}x + \delta|x| \leq 0. \end{aligned} \tag{LRC}$$

As already investigated in Example 2.48, the feasible set mapping and the optimal set mapping are given by

$$\mathcal{F}(u) = \begin{cases} [0,1] & \text{if } u < 0 \\ [-1,1] & \text{if } u = 0 \\ [-1,0] & \text{if } u > 0 \end{cases} \quad \text{and} \quad \mathcal{F}^*(u) = \begin{cases} \{1\} & \text{if } u < 0 \\ \{1\} & \text{if } u = 0 \\ \{0\} & \text{if } u > 0. \end{cases}$$

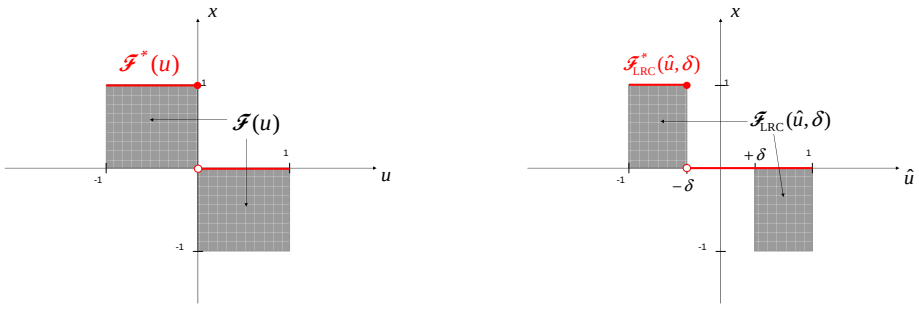
Note that the feasible set mapping $\mathcal{F}$ is only Hausdorff upper semicontinuous at $\bar{u} = 0$, but not Hausdorff lower semicontinuous (there does not exist a Slater point). For the robust problem (LRC), the respective mappings are

$$\mathcal{F}_{LRC}(\hat{u},\delta) = \begin{cases} [0,1] & \text{if } \hat{u} \leq -\delta \\ \{0\} & \text{if } -\delta < \hat{u} < \delta \\ [-1,0] & \text{if } \hat{u} \geq \delta \end{cases} \quad \text{and} \quad \mathcal{F}^*_{LRC}(\hat{u},\delta) = \begin{cases} \{1\} & \text{if } \hat{u} \leq -\delta \\ \{0\} & \text{if } -\delta < \hat{u} < \delta \\ \{0\} & \text{if } \hat{u} \geq \delta. \end{cases}$$

We again consider the particular point $\bar{u} = 0$ and let $\delta \to 0$. Then it holds that

$$\{0\} = \mathcal{F}^*_{LRC}(0,\delta) \nrightarrow \mathcal{F}^*_{LRC}(0,0) = \{1\} = \mathcal{F}^*(0),$$

i.e. the sequence of robust optimal solutions does not converge to an optimal solution of the original problem. Figure 3.2 illustrates the set of feasible and optimal solutions in both the original problem (Figure 3.2(a)) and the local robust counterpart problem (Figure 3.2(b)).



(a) original problem (b) robust problem

Figure 3.2: Illustration of the sets of feasible and optimal solutions of the original and the robust program of Example 3.24.

At the end of this section, we again summarize in Figure 3.3 the stability results in a diagram analogous to Figure 2.8.

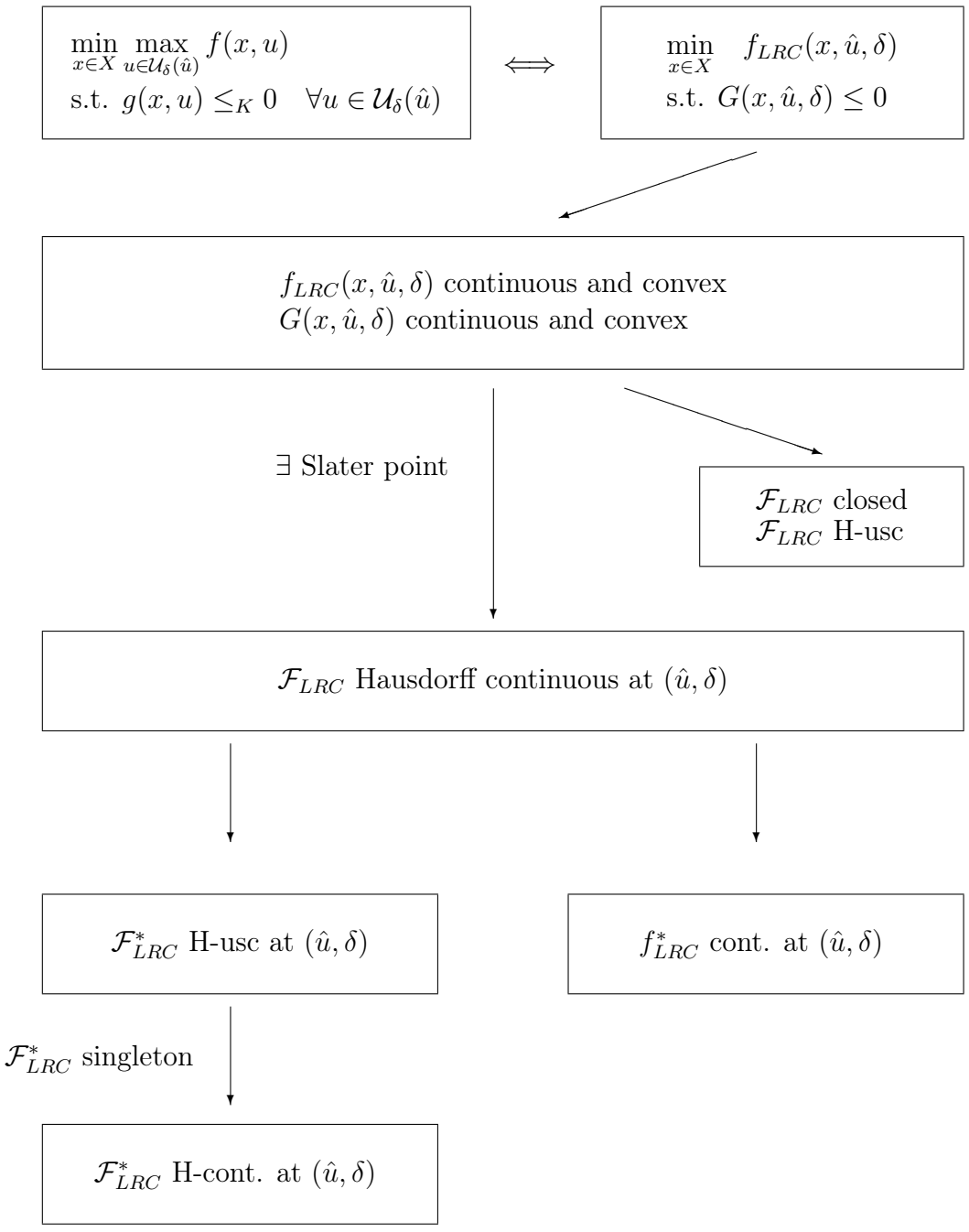


Figure 3.3: Illustration of the continuity results for $(LRC_{\hat{u},\delta})$.

3.3 Influence of the shape of the uncertainty set

In this section we investigate the influence a particular shape of the uncertainty set can have on the continuity properties of the (set of) optimal solutions. We will prove that in many practical applications using an uncertainty set with an ellipsoidal shape leads to a certain structure in the set of optimal solutions and rather often even to a unique solution, thus as well continuity. But first, we illustrate by a very simple one-dimensional example that the the (local) robust counterpart approach does not necessarily lead to a continuous solution in general.

Example 3.25. *Consider the optimization problem of Example 3.24:*

$$\begin{aligned} \min_{x\in[-1,1]} \quad & (x-1)^2 \\ s.t. \quad & ux \leq 0 \end{aligned} \tag{P}$$

with $u \in \mathcal{U} = [-1,1]$. Letting the local uncertainty set be described by $\mathcal{U}_\delta(\hat{u}) = [\hat{u}-\delta, \hat{u}+\delta]$ with $\delta > 0$, the local robust counterpart reformulates to

$$\begin{aligned} \min_{x\in[-1,1]} \quad & (x-1)^2 \\ s.t. \quad & \hat{u}x + \delta|x| \leq 0. \end{aligned} \tag{LRC}$$

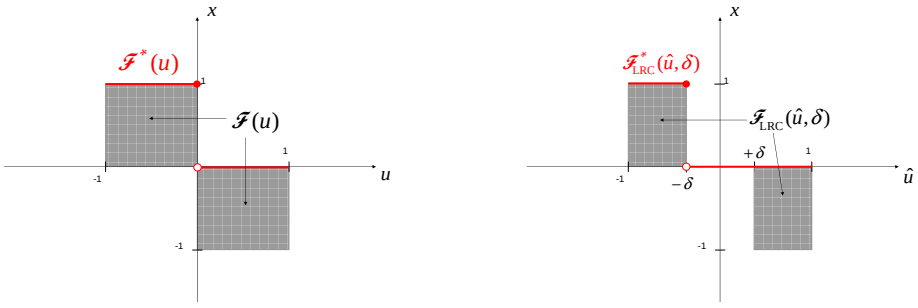
In Example 2.48 we have already seen that both the feasible set mapping $\mathcal{F}$ and the optimal set mapping $\mathcal{F}^*$ of the original problem (P) are not continuous is $u = 0$. From Examle 3.24, we can recall the feasible set mapping $\mathcal{F}_{LRC}$ and the optimal set mapping $\mathcal{F}^*_{LRC}$:

$$\mathcal{F}_{LRC}(\hat{u},\delta) = \begin{cases} [0,1] & \text{if } \hat{u} \leq -\delta \\ 0 & \text{if } -\delta < \hat{u} < \delta \\ [-1,0] & \text{if } \hat{u} \geq \delta \end{cases}$$

and

$$\mathcal{F}^*_{LRC}(\hat{u},\delta) = \begin{cases} 1 & \text{if } \hat{u} \leq -\delta \\ 0 & \text{if } -\delta < \hat{u} < \delta \\ 0 & \text{if } \hat{u} \geq \delta \end{cases}$$

which are both still not continuous for all parameter values. Hence, the local robust counterpart approach does not generally help to change the original program to a $\mathcal{U}$-stable one even though in this example it creates continuity at the point $u = 0$. But the discontinuity in the optimal set mapping is not eliminated completely, the critical position is only relocated from $u = 0$ to the point $\hat{u} = -\delta$. For illustration we again show in Figure 3.4 the sets of feasible and optimal solutions for both the original and the robust problem.



(a) original problem (b) robust problem

Figure 3.4: Illustration of the sets of feasible and optimal solutions of the original and the robust program of Example 3.25.

Thus, this small example already shows that the robust counterpart approach cannot be used as a general method to make a program $\mathcal{U}$-stable. It seems that it stabilizes the solution locally around the particular parameter $\hat{u}$, but not on the whole uncertainty set $\mathcal{U}$.

Despite the drawback illustrated in the small example, the (local) robust counterpart approach is a very useful method. In this preceding example we could not exploit all the possibilities of the approach because of the one-dimensionality which reduces any uncertainty set around a given parameter $\hat{u}$ to an interval. In the following we analyze the influence of different choices of the *shape* of the uncertainty set $\mathcal{U}$.

The first example illustrates in case of a linear function how the expression "$\max_{u\in\mathcal{U}_\delta(\hat{u})}$" can be reformulated for two explicitly given uncertainty sets and thus leads to a tractable optimization problem. Afterwards we will deal with the solution and stability considerations of the particular problem and its associated robust counterparts using different shapes of uncertainty.

Example 3.26. *In this example we consider the two most intuitive uncertainty sets: interval or box uncertainty and ellipsoidal uncertainty. We choose for the illustration the simple program*

$$\min_{x\in X} \quad -x^T u$$

with $X \subset \{x \in \mathbb{R}^n \mid x \geq 0\}$ *being a non-empty, convex and compact set. The corresponding local robust counterpart program is thus generally given by*

$$\min_{x\in X} \quad \max_{u\in\mathcal{U}_\delta(\hat{u})} -x^T u.$$

First we use an interval or box uncertainty set $\mathcal{U}_{\delta,box}(\hat{u})$ *of size* $\delta > 0$ *around a given parameter* $\hat{u} \in \mathcal{U}$*. Here, each component* u_i *of the parameter vector can*

vary independently within its interval around $\hat{u}_i$. The set is then given by

$$\mathcal{U}_{\delta,box}(\hat{u}) = \{u \in \mathcal{U} \mid u = \hat{u} + \delta w, w \in [-1,1]^n\}.$$

The worst case solution of the robust counterpart program with such an uncertainty set is obviously attained when the explicit parameter $\hat{u} - \delta\mathbf{1}$ is used for the optimization. Therefore, the local robust counterpart to the above program is then given by:

$$\begin{aligned} &\min_{x \in X} \quad -x^T(\hat{u} - \delta\mathbf{1}) \\ =\ &\min_{x \in X} \quad -x^T\hat{u} + \delta x^T\mathbf{1} \end{aligned}$$

which is now rather easily solvable.

Second, we consider an ellipsoidal uncertainty set $\mathcal{U}_{\delta,ell}(\hat{u})$ of size $\delta > 0$ around a given parameter $\hat{u} \in \mathcal{U}$. The matrix Σ describing the shape of the ellipsoid is assumed to be symmetric and positive definite. Thus,

$$\begin{aligned} \mathcal{U}_{\delta,ell}(\hat{u}) &= \left\{u \in \mathcal{U} \mid (u-\hat{u})^T\Sigma^{-1}(u-\hat{u}) \leq \delta^2\right\} \\ &= \left\{u \in \mathcal{U} \mid u = \hat{u} + \delta\,\Sigma^{\frac{1}{2}}w, \|w\| \leq 1\right\}. \end{aligned}$$

The equivalence of these sets is shown in Appendix E. Using this particular uncertainty set $\mathcal{U}_{\delta,ell}(\hat{u})$, we can reformulate the program as follows:

$$\begin{aligned} \min_{x \in X} \quad &\max_{u \in \mathcal{U}_{\delta,\,ell}(\hat{u})} \quad -x^Tu \\ &= \min_{x \in X} \quad \max_{\|w\| \leq 1} \quad -x^T\hat{u} - \delta x^T\Sigma^{\frac{1}{2}}w \\ &= \min_{x \in X} \left(-x^T\hat{u} + \delta \max_{\|w\| \leq 1} -x^T\Sigma^{\frac{1}{2}}w\right) \end{aligned}$$

and since the negative scalar product of $(\Sigma^{\frac{1}{2}}x)$ and w is largest for $w^* = -\dfrac{\Sigma^{\frac{1}{2}}x}{\|\Sigma^{\frac{1}{2}}x\|}$ this gives

$$\begin{aligned} &= \min_{x \in X} \quad -x^T\hat{u} + \delta\, x^T\Sigma^{\frac{1}{2}}\,\frac{\Sigma^{\frac{1}{2}}x}{\|\Sigma^{\frac{1}{2}}x\|} \\ &= \min_{x \in X} \quad -x^T\hat{u} + \delta\,\|\Sigma^{\frac{1}{2}}x\|. \end{aligned}$$

In addition to the reformulation of the robust problem, the worst case parameter can also be stated explicitly:

$$u_{wc} = \hat{u} - \delta\,\frac{\Sigma x}{\|\Sigma^{\frac{1}{2}}x\|}.$$

After having seen the reformulation of the local robust counterpart program given two explicit uncertainty sets, we now investigate the stability of these robustified programs. We have already analyzed the same problem in Section 2.3.6 and found that the optimal solution is discontinuous at the point where the components of u are equal. In the following example we now apply the particular shapes of a square and a circular[4] uncertainty set and examine the effects thereof on the stability of the robustified program.

Example 3.27. *Consider the following optimization problem:*

$$\min_{x \in X} \quad -x^T u \tag{P}$$

with $X = \{x \in \mathbb{R}^2 \mid x \geq 0, x^T \mathbf{1} = 1\}$ *and* $u \in \mathcal{U} \subset \mathbb{R}^2$. *Note that this particular problem was already investigated in Section 2.3.6 and that it is a special case of the n-dimensional program from the previous example. Recall from Section 2.3.6 the following established facts about (P):*

- *The extreme value function is continuous on* $\mathcal{U}$.
- *The optimal set mapping is discontinuous at the points* u *with* $u_1 = u_2$, *thus, there does not exist a continuous selection function within the set of optimal solutions.*
- *The* ε*-optimal set mapping is Hausdorff lower semicontinuous and thus there exists a continuous selection function within* $\mathcal{F}^*_\varepsilon$.

In this example we are interested in improving the result about the optimal set mapping. We apply the local robust counterpart idea with different uncertainty sets to examine the influence of the particular shape on the set of optimal solutions in each case.

Before starting the calculations, we recall the used notation to avoid ambiguity. We will consider the (large) uncertainty set $\mathcal{U}$ *to be centered at* u_0, *once assuming the shape of a square and once the shape of a circle, representing the two cases of interval and ellipsoidal uncertainty. The local robust counterpart will then be formulated around the particular parameter* $\hat{u} \in \mathcal{U}$ *where we assume the center* $\hat{u}$ *and the size* δ *to be chosen in such a way that* $\mathcal{U}_\delta(\hat{u}) \subset \mathcal{U}$. *Furthermore, the local uncertainty set is supposed to have the same shape as* $\mathcal{U}$. *This general link of* $\mathcal{U}$ *and* $\mathcal{U}_\delta(\hat{u})$ *was already shown in Figure 3.1. The following Figure 3.5 illustrates the two particular cases we want to investigate in this example.*

The explicit formulations of these local uncertainty sets are given by

$$\mathcal{U}_{\delta,box}(\hat{u}) = \left\{u \in \mathcal{U} \mid u = \hat{u} + \delta w, w \in [-1,1]^2\right\}$$

[4]The same qualitative results hold if a general ellipse is chosen instead of a circle, but for simplicity of the explicit calculations we used the special case of $\Sigma = I$, the identity matrix.

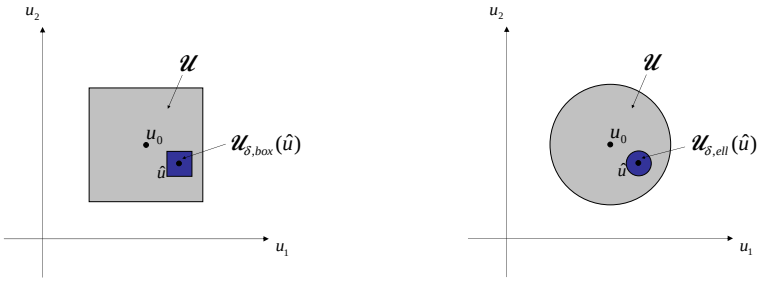


Figure 3.5: Illustration of the two uncertainty set $\mathcal{U}_{\text{box}}$ and $\mathcal{U}_{\text{ell}}$.

and

$$\mathcal{U}_{\delta,ell}(\hat{u}) = \{u = \hat{u} + \delta w \mid \|w\|_2 \leq 1\}$$

and the general representation of the local robust counterpart program for this example is the following:

$$\min_{x \in X} \max_{u \in \mathcal{U}_\delta(\hat{u})} \quad -x^T u \tag{LRC}$$

with $X = \{x \in \mathbb{R}^2 \mid x \geq 0,\ x^T\mathbf{1} = 1\}$. *This problem (LRC) will now be reformulated according to the chosen shape of* $\mathcal{U}_\delta(\hat{u})$.

First, we consider the program together with interval uncertainty. In Example 3.26 we have already seen that the (local) robust counterpart program in our particular problem can be reformulated simply by inserting the worst case feasible parameter choice. Thus, the final problem we have to solve in this case, is:

$$\min_{x \in X} \quad -x^T(\hat{u} - \delta\mathbf{1}).$$

But this is exactly the same type of problem as (P) itself, just with a different parameter $\hat{u} - \delta\mathbf{1}$ *instead of* u. *Hence, we already know the solutions for the extreme value function* $f^*_{LRC}(\hat{u}, \delta)$ *and the set of optimal solutions* $\mathcal{F}^*_{LRC}(\hat{u}, \delta)$[5]*:*

- *The extreme value function* $f^*_{LRC}(\hat{u}, \delta)$ *with interval uncertainty is*

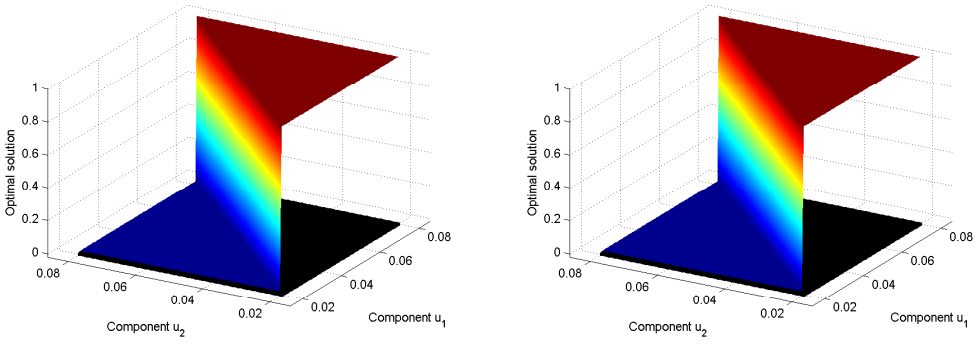
$$\begin{aligned} f^*_{LRC}(\hat{u}, \delta) &= \min_{x \in X}\{-x^T(\hat{u} - \delta\mathbf{1})\} \\ &= f^*(\hat{u} - \delta\mathbf{1}) \quad \textit{with } f^* \textit{ as in Section 2.3.6 .} \end{aligned}$$

[5]Note that we are interested in continuity of the optimal set mapping $\mathcal{F}^*_{LRC}$, since this was not given in the original problem (P). As we could already determine a continuous selection function within the ε-optimal set mapping in (P), there is no need for explicitly investigating the ε-optimal set mapping $\mathcal{F}^*_{\varepsilon,LRC}(\hat{u}, \delta)$.

- *The associated optimal set mapping is analogously given by*

$$\mathcal{F}^*_{LRC}(\hat{u}, \delta) = \mathcal{F}^*(\hat{u} - \delta\mathbf{1}) \quad \textit{with } \mathcal{F}^* \textit{ as in Section 2.3.6 .}$$

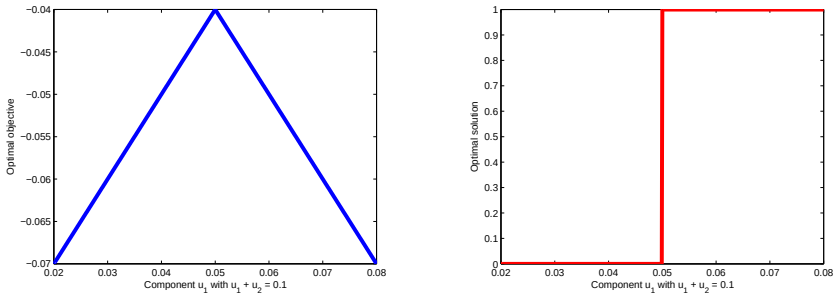
Since in this case of square uncertainty we can simply use the results from Section 2.3.6, we already know that the optimal set mapping $\mathcal{F}^*_{LRC}$ is not continuous. The graphical illustration of the optimal solution (represented by the first component x_1, the second component is simply given by $1 - x_1$) for various values of u or $\hat{u}$, respectively, thus is identical and shown again in Figure 3.6.



(a) optimal solution $\mathcal{F}^*(u)$ (b) optimal solution $\mathcal{F}^*_{LRC}(u)$

Figure 3.6: Illustration of the optimal solutions $\mathcal{F}^*$ and $\mathcal{F}^*_{LRC}$ in Example 3.27 using a box uncertainty set.

Figure 3.7(b) shows the selected view on the optimal solution along the diagonal, i.e. along the line where the sum of the two components is constant. The graph shows the optimal weight in asset 1, plotted against the value of the respective first component of the vector u while it holds that $u_1 + u_2 = 0.1$ to represent the diagonal in the above Figure 3.6.



(a) objective $f^*_{LRC}(u)$ (b) optimal solution $\mathcal{F}^*_{LRC}(u)$

Figure 3.7: Illustration of f^*_{LRC} and $\mathcal{F}^*_{LRC}$ along the diagonal using a box uncertainty set.

Note that the midpoint $u_1 = 0.05$ *denotes the vector with equal components, i.e.* $u_1 - u_2 = 0$, *and to the sides the expression* $\Delta u = |u_1 - u_2|$ *increases. Figure 3.7(a) illustrates the associated extreme value function given by* $f^*_{LRC}(u) = \min\{-(u_1 - \delta), -(u_2 - \delta)\}$.

We next examine problem (LRC) using the circular uncertainty set $\mathcal{U}_{\delta,ell}(\hat{u})$. *The robust counterpart reformulation in this case yields*

$$\min_{x \in X} \quad -x^T \hat{u} + \delta \|x\|$$

according to Example 3.26 with $\Sigma = I$.

Since we are working in the two dimensional space, it is possible to give explicit solutions[6] *to the functions* $f^*_{LRC}(\hat{u}, \delta)$ *and* $\mathcal{F}^*_{LRC}(\hat{u}, \delta)$ *we are interested in.*

- *The extreme value function* $f^*_{LRC}(\hat{u}, \delta)$ *with circular uncertainty is given by*

$$f^*_{LRC}(\hat{u}, \delta) =$$

$$= \begin{cases} -\hat{u}_1 + \delta & \text{if } \hat{u}_1 \geq \hat{u}_2 + \delta \\ -\frac{1}{2}(\hat{u}_1 + \hat{u}_2) + \frac{1}{2}\sqrt{2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2} & \text{if } \hat{u}_2 + \delta > \hat{u}_1 > \hat{u}_2 \\ -\hat{u}_1 + \frac{\delta}{\sqrt{2}} & \text{if } \hat{u}_1 = \hat{u}_2 \\ -\frac{1}{2}(\hat{u}_1 + \hat{u}_2) + \frac{1}{2}\sqrt{2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2} & \text{if } \hat{u}_1 < \hat{u}_2 < \hat{u}_1 + \delta \\ -\hat{u}_2 + \delta & \text{if } \hat{u}_1 + \delta \leq \hat{u}_2. \end{cases}$$

- *The associated optimal set mapping is*

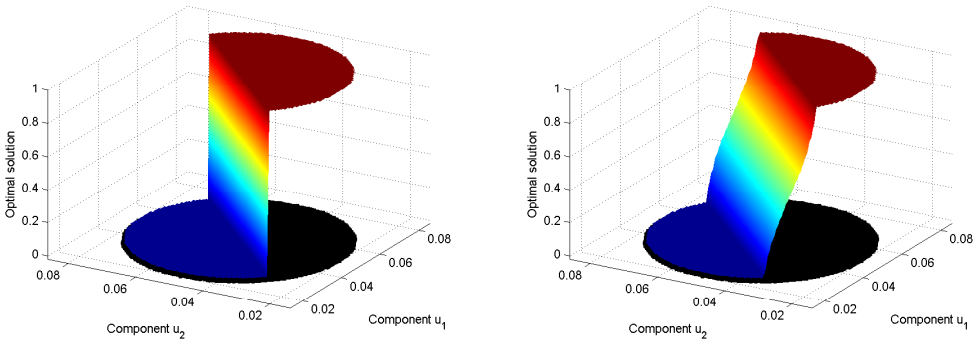
$$\mathcal{F}^*_{LRC}(\hat{u}, \delta) =$$

$$= \begin{cases} \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \text{if } \hat{u}_1 \geq \hat{u}_2 + \delta \\ \begin{pmatrix} \min\left\{\frac{1}{2} + \frac{|\hat{u}_1 - \hat{u}_2|}{2\sqrt{2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2}}; 1\right\} \\ \max\left\{\frac{1}{2} - \frac{|\hat{u}_1 - \hat{u}_2|}{2\sqrt{2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2}}; 0\right\} \end{pmatrix} & \text{if } \hat{u}_2 + \delta > \hat{u}_1 > \hat{u}_2 \\ \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} & \text{if } \hat{u}_1 = \hat{u}_2 \\ \begin{pmatrix} \max\left\{\frac{1}{2} - \frac{|\hat{u}_1 - \hat{u}_2|}{2\sqrt{2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2}}; 0\right\} \\ \min\left\{\frac{1}{2} + \frac{|\hat{u}_1 - \hat{u}_2|}{2\sqrt{2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2}}; 1\right\} \end{pmatrix} & \text{if } \hat{u}_1 < \hat{u}_2 < \hat{u}_1 + \delta \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \text{if } \hat{u}_1 + \delta \leq \hat{u}_2. \end{cases}$$

[6] For the detailed calculations see Appendix G.

Thus in this case, the set of optimal solutions $\mathcal{F}^*_{LRC}(\hat{u}, \delta)$ is a singleton for any parameter choice $\hat{u}$, and this piecewise defined function is continuous in $\hat{u}$.

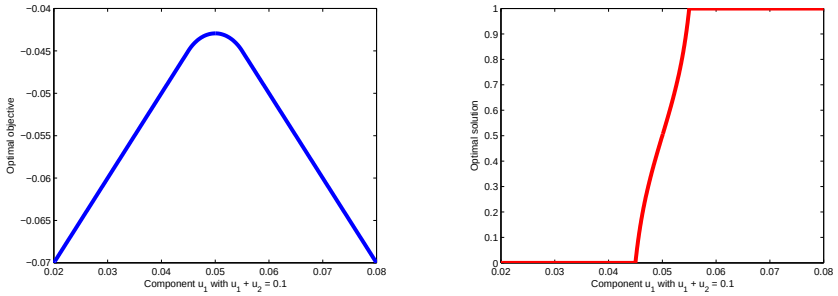
Figure 3.8(b) illustrates the optimal solution $\mathcal{F}^*_{LRC}(\hat{u}, \delta)$, represented again by the first component. For comparison, we included as well the plot of the optimal solution of the orginal problem (P) in Figure 3.8(a), as already shown in Section 2.3.6. It can nicely be seen that the line where in the previous example the "jump" or discontinuity has occured is now smoothed. At this line, the optimal solution is always $(0.5, 0.5)^T$ and this solution changes continuously to one of the extremes $(1, 0)^T$ or $(0, 1)^T$, respectively, as the difference between the components of $\hat{u}$ increases.



(a) optimal solution $\mathcal{F}^*(u)$ (b) optimal solution $\mathcal{F}^*_{LRC}(u)$

Figure 3.8: Illustration of the optimal solutions $\mathcal{F}^*$ and $\mathcal{F}^*_{LRC}$ in Example 3.27 using an ellipsoidal uncertainty set.

In this case of robustification using a circular uncertainty set, we also show in Figure 3.9 the results along the diagonal.



(a) objective $f^*_{LRC}(u)$ (b) optimal solution $\mathcal{F}^*_{LRC}(u)$

Figure 3.9: Illustration of f^*_{LRC} and $\mathcal{F}^*_{LRC}$ along the diagonal using a circular uncertainty set.

As can be observed from Figure 3.9(b), the optimal set mapping (again represented by the weight in asset 1) is now continuous in contrast to the robustification

using a box uncertainty set as shown in Figure 3.7(b). The critical point at $\hat{u}$ in the original problem is eliminated, but at the boundary of the uncertainty set, i.e. at $\hat{u} \pm \delta$, "corners" are introduced. Furthermore, the extreme value function (Figure 3.9(a)) is also smoothed in the δ-neighborhood around the point $\hat{u}$ where $u_1 = u_2$.

In the preceding examples we have seen that the robust counterpart in general does not yet guarantee $\mathcal{U}$-stability of the program – it depends on the chosen shape of the uncertainty set $\mathcal{U}$. Interval uncertainty only shifted the point with the discontinuity, thus not improving the situation as a whole, but an ellipsoidal uncertainty set seems promising.

In the following we now analyze and state more theoretically in which cases, i.e. under which (special) conditions the robust counterpart leads to the desired result of making a program $\mathcal{U}$-stable.

Theorem 3.28 (Benefits of robustification). *Consider program (GCP_u) and assume the objective function f to have the form $f(x,u) = f_0(x) + (Ax)^T u$ with $f_0 : \mathbb{R}^n \to \mathbb{R}$ being twice differentiable and convex and $A \in \mathbb{R}^{d \times n}$. Furthermore, let the local uncertainty set $\mathcal{U}_\delta(\hat{u})$ have ellipsoidal shape, i.e.*

$$\mathcal{U}_\delta(\hat{u}) = \{u \in \mathbb{R}^d \mid u = \hat{u} + \delta H w, \|w\| \leq 1\}$$

with $H \in \mathbb{R}^{d \times d}$ symmetric and positive definite.
Then there exists an $x^* \in \mathcal{F}_{\mathcal{U}_\delta(\hat{u})}$[7] such that the following holds for the optimal solution set $\mathcal{F}^*_{LRC}(\hat{u}, \delta)$ of the local robust counterpart:

(i) $\mathcal{F}^*_{LRC}(\hat{u}, \delta) = \{x^*\}$, *i.e. a singleton, or*

(ii) $\mathcal{F}^*_{LRC}(\hat{u}, \delta) = \{y^* \mid y^* = \lambda x^* + z, \lambda \in \mathbb{R}, z \in \ker(HA)\} \cap \mathcal{F}_{\mathcal{U}_\delta(\hat{u})}$.

with $\ker(B) := \{x \in \mathbb{R}^n \mid Bx = 0\}$.

Remark 3.29.

(i) Note that in this case of H being positive definite, $\ker(HA) = \ker(A)$. *We will nevertheless use the notation* $\ker(HA)$ *to indicate the dependence on the particular form of the ellipsoidal uncertainty set.*

(ii) Furthermore, the requirement of H being positive definite is not a restriction, as the dimension d of the uncertain parameters can without loss of generality be chosen such that H is positive definite – by possibly reducing the uncertain vector to those components that really are exposed to uncertainty.

[7] Recall that $\mathcal{F}_{\mathcal{U}_\delta(\hat{u})} = \bigcap_{u \in \mathcal{U}_\delta(\hat{u})} \mathcal{F}(u)$.

In the proof of Theorem 3.28 we will need the result of the following lemma.

Lemma 3.30. *Let $v \in \mathbb{R}^d$ with $\|v\| = 1$, $l \in \mathbb{R}^d, l \neq 0$ and let*

$$l^T[I - vv^T]l = 0.$$

Then it holds that the vector l must be a multiple of v, i.e. there exists a $k \in \mathbb{R}$ such that $l = k\,v$.

Proof. First of all note that the matrix $[I - vv^T] \in \mathbb{R}^{d\times d}$ is positive semidefinite, since

$$\begin{aligned} w^T[I - vv^T]w &= \|w\|^2 - (v^Tw)^T(v^Tw) = \|w\|^2 - \|v^Tw\|^2 \\ &\overset{\text{Cauchy-Schwarz}}{\geq} \|w\|^2 - \underbrace{\|v\|^2}_{=1}\cdot\|w\|^2 = 0 \end{aligned}$$

with $w \in \mathbb{R}^d$ being an arbitrary vector. It is not positive definite as the vector $l \neq 0$ fulfills the equation $l^T[I - vv^T]l = 0$. Hence, the matrix $[I - vv^T]$ must have a zero eigenvalue with v being the corresponding normed eigenvector, since $[I - vv^T]v = 0 = 0 \cdot v$. As the rank of the matrix is at least $d - 1$ (due to subtraction of a dyadic product, a rank 1 matrix), we can conclude from the dimension formula that the kernel of the matrix (i.e. the space of eigenvectors to the eigenvalue 0) has the dimension 1.

Furthermore, as the vector l is an eigenvector to the eigenvalue zero as well[8], l must be a multiple of v, i.e. there exists a $k > 0$ such that $l = k\,v$. □

Proof of Theorem 3.28. Applying the robust counterpart approach and using the reformulation as in Example 3.26, the robust objective function is given by

$$f_{LRC}(x, \hat{u}, \delta) = f_0(x) + (Ax)^T\hat{u} + \delta\|HAx\|$$

which is again a convex function as $\|.\|$ is convex. For later reference, we provide the first two derivatives thereof with respect to the variable x:

$$\begin{aligned} f'_{LRC}(x, \hat{u}, \delta) &= f'_0(x) + A^T\hat{u} + \delta\frac{(HA)^THAx}{\|HAx\|}, \\ f''_{LRC}(x, \hat{u}, \delta) &= f''_0(x) + \delta\frac{1}{\|HAx\|}(HA)^T\left[I - \frac{HAx}{\|HAx\|}\frac{(HAx)^T}{\|HAx\|}\right](HA). \end{aligned}$$

Let $x^* \in \mathcal{F}^*_{LRC}(\hat{u}, \delta)$. Note that $\mathcal{F}^*_{LRC}(\hat{u}, \delta) \neq \emptyset$ since the feasibility set is non-empty (see Assumption 2.17) and bounded. If x^* is the only solution of the local robust counterpart program, we are done. Otherwise, we consider two cases:

[8] Note that it holds $l^T[I - vv^T]l = \left([I - vv^T]^{\frac{1}{2}}l\right)^T\left([I - vv^T]^{\frac{1}{2}}l\right) = 0$, i.e. $[I - vv^T]^{\frac{1}{2}}l = 0$ implying that l is an eigenvector to the eigenvalue zero for the matrix $[I - vv^T]^{\frac{1}{2}}$, hence for $[I - vv^T]$.

- All optimal solutions lie within $\ker(HA)$, i.e. $\mathcal{F}^*_{LRC}(\hat{u},\delta) \subset \ker(HA)$. Let $y^* \in \mathcal{F}^*_{LRC}(\hat{u},\delta), y^* \neq x^*$. Then the representation of y^* in (ii) trivially holds for $z := -\lambda x^* + y^* \in \ker(HA)$, $\lambda \in \mathbb{R}$.

- There exists at least one optimal solution not lying within $\ker(HA)$. Without loss of generality, let $x^* \in \mathcal{F}^*_{LRC}(\hat{u},\delta), x^* \notin \ker(HA)$. Let $y^* \neq x^*$ be an arbitrary further optimal solution (within $\ker(HA)$ or not) and define $h := y^* - x^*$. Since the set of optimal solutions of a convex problem is a convex set (see e.g. Jahn [42], Theorem 2.14), all the points $z_\alpha := x^* + \alpha h$ for $\alpha \in [0,1]$ are optimal solutions of $(LRC_{\hat{u},\delta})$ as well. Therefore, we have

$$f^*_{LRC}(\hat{u},\delta) = f_{LRC}(x^*,\hat{u},\delta) = f_{LRC}(z_\alpha,\hat{u},\delta)$$

for all $\alpha \in [0,1]$. Taylor expansion of $f_{LRC}(z_\alpha,\hat{u},\delta)$ at x^* yields

$$\begin{aligned} f_{LRC}(z_\alpha,\hat{u},\delta) = \; & f_{LRC}(x^*,\hat{u},\delta) + (f'_{LRC}(x^*,\hat{u},\delta))^T \cdot \alpha h \\ & + \frac{1}{2}(\alpha h)^T f''_{LRC}(x^*,\hat{u},\delta)(\alpha h) + o(\alpha^2) \end{aligned}$$

and thus

$$0 = (f'_{LRC}(x^*,\hat{u},\delta))^T \cdot \alpha h + \frac{1}{2}(\alpha h)^T f''_{LRC}(x^*,\hat{u},\delta)(\alpha h) + o(\alpha^2). \tag{3.1}$$

Dividing Equation (3.1) by $\alpha > 0$ and taking the limit $\alpha \to 0$ gives

$$\begin{aligned} 0 &= (f'_{LRC}(x^*,\hat{u},\delta))^T h + \lim_{\alpha\to 0}\frac{1}{2}\alpha h^T f''_{LRC}(x^*,\hat{u},\delta)h + \lim_{\alpha\to 0}\frac{o(\alpha^2)}{\alpha} \\ &= (f'_{LRC}(x^*,\hat{u},\delta))^T h. \end{aligned}$$

Using this result in the above Equation (3.1), dividing the remaining terms again by α and taking the limit thus yields

$$\begin{aligned} 0 &= \frac{1}{2}h^T f''_{LRC}(x^*,\hat{u},\delta)\, h + \lim_{\alpha\to 0}\frac{o(\alpha^2)}{\alpha^2} \\ &= \frac{1}{2}h^T f''_{LRC}(x^*,\hat{u},\delta)\, h \qquad (3.2) \\ &= \frac{1}{2}h^T \left(f_0''(x^*) + \frac{\delta}{\|HAx^*\|}(HA)^T \left[I - \frac{HAx^*}{\|HAx^*\|}\frac{(HAx^*)^T}{\|HAx^*\|} \right](HA) \right) h. \qquad (3.3) \end{aligned}$$

Since both $f_0(x)$ and the norm function $\|.\|$ are convex, the respective Hessian matrices $f_0''(x^*)$ and $I - \dfrac{HAx^*}{\|HAx^*\|}\dfrac{(HAx^*)^T}{\|HAx^*\|}$ are positive semidefinite.

Thus it holds for all $h \in \mathbb{R}^n$ that

$$0 \leq h^T f_0''(x^*)h \quad \text{and}$$
$$0 \leq \frac{\delta}{\|HAx^*\|} h^T (HA)^T \left[I - \frac{HAx^*}{\|HAx^*\|} \frac{(HAx^*)^T}{\|HAx^*\|} \right] (HA)\, h.$$

Hence, together with Equation (3.2) we conclude that it must hold

$$0 = h^T f_0''(x^*)h \quad \text{and} \tag{3.4}$$
$$0 = \frac{\delta}{\|HAx^*\|} h^T (HA)^T \left[I - \frac{HAx^*}{\|HAx^*\|} \frac{(HAx^*)^T}{\|HAx^*\|} \right] (HA)\, h. \tag{3.5}$$

Focusing on Equation (3.5), we again distinguish two cases:

1. $h \in \ker(HA)$. Then we are done as $y^* = x^* + h$.
2. $h \notin \ker(HA)$, i.e. $l := (HA)h \neq 0, l \in \mathbb{R}^d$. With $v = \frac{HAx^*}{\|HAx^*\|}$, $\|v\| = 1$, Equation (3.5) can be written in simplified form as $l^T[I - vv^T]l = 0$. Using Lemma 3.30 we can thus conclude that the vector l is a multiple of v, i.e. there exists a $k \in \mathbb{R}$ such that

$$\begin{aligned}(HA)h = l &= k\, v \\ &= k\, \frac{HAx^*}{\|HAx^*\|}\end{aligned}$$

and thus

$$\begin{aligned}HAy^* &= HAx^* + HAh \\ &= HAx^* + \frac{k}{\|HAx^*\|} HAx^* \\ &= \left(\frac{k}{\|HAx^*\|} + 1 \right) HAx^*\end{aligned}$$

which yields

$$y^* = \underbrace{\left(\frac{k}{\|HAx^*\|} + 1 \right)}_{=:\lambda} x^* + z, z \in \ker(HA).$$

In any case, intersecting the solution set with $\mathcal{F}_{\mathcal{U}_\delta(\hat{u})}$ concludes the proof.

Note that Equation (3.4) additionally restricts the set of optimal solutions, but as $f_0(x)$ was an arbitrary convex function, these conditions are not generally expressable but depend on the particular function. □

Corollary 3.31. *Let the assumptions of Theorem 3.28 hold. Furthermore, suppose that the matrix A has full column rank n (i.e. especially implying $d \geq n$). Then the optimal solution $\mathcal{F}^*_{LRC}(\hat{u}, \delta)$ of the local robust counterpart is either a singleton or contains only linearly dependent solutions, i.e. there exists an $x^* \in \mathcal{F}_{\mathcal{U}_\delta(\hat{u})}$ such that*

$$\mathcal{F}^*_{LRC}(\hat{u}, \delta) = \{y^* \mid y^* = \lambda\ x^*, \lambda \in [\lambda_l, \lambda_u] \subset \mathbb{R}\}\,.$$

Proof. Given that $\operatorname{rank}(A) = n$ and H positive definite, thus non-singular, the matrix product HA also has full rank n, i.e. $\ker(HA) = \{0\}$. Using this fact and the result of Theorem 3.28, we hence get

$$\mathcal{F}^*_{LRC}(\hat{u}, \delta) = \{y^* \mid y^* = \lambda\ x^*, \lambda \in \mathbb{R}\} \cap \mathcal{F}_{\mathcal{U}_\delta(\hat{u})}.$$

Since X is a compact set, this equation is equivalent to restricting λ to a compact interval of $\mathbb{R}$, i.e.

$$\mathcal{F}^*_{LRC}(\hat{u}, \delta) = \{y^* \mid y^* = \lambda\ x^*, \lambda \in [\lambda_l, \lambda_u]\}\,. \qquad \square$$

Corollary 3.32. *Let the assumptions of Corollary 3.31 hold, and assume further that $X \subset \{x \in \mathbb{R}^n \mid a^T x = b\}$. Then the optimal solution of the local robust counterpart problem is unique.*

Proof. Direct consequence of Corollary 3.31 since the constraint $a^T x = b$ excludes multiples of x^*. $\square$

Remark 3.33. *Corollary 3.32 hence gives that additionally imposing certain non-parallel constraints (non-parallel to the vector x^*) yields a unique optimal solution of the robust program. Furthermore, by Proposition 3.21, part (v), we know that this unique solution is also stable, i.e. continuous in $(\hat{u}, \delta)$ under the prerequisite of a continuous feasibility set.*

In portfolio optimization problems the set of constraints usually contains the equation $x^T\mathbf{1} = 1$ which defines the vector x to represent a portfolio. Hence, in all the portfolio applications we will obtain a unique optimal solution of the robust problem formulation when using an ellipsoidal uncertainty set.

Figure 3.10 illustrates the result of Corollary 3.31 and the further implication of non-parallel constraints.

The result presented in Theorem 3.28 and in particular the consequences thereof as stated in Corollary 3.32 will be applied in many situations in the second part of this dissertation. Due to this result we will first of all create only ellipsoidal uncertainty sets for the practical applications, and second, we will be able to proof uniqueness of the optimal solution of the robust portfolio optimization problem.

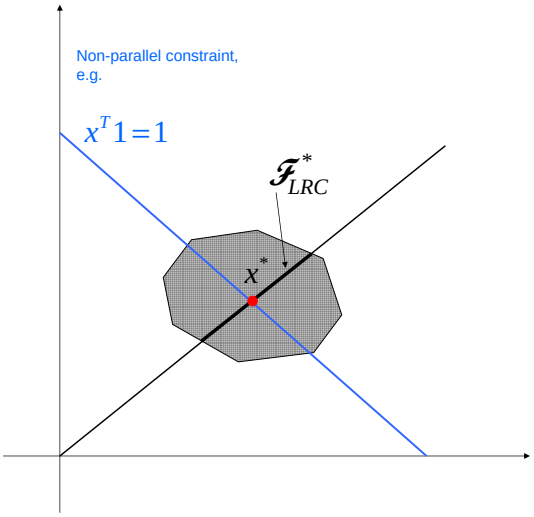


Figure 3.10: Illustration of the benefits of robustification as described in Corollaries 3.31 and 3.32.

At the end of this section we want to summarize the bottom line of our results. We have seen that one of the most natural choices for the shape of the local uncertainty set – an ellipsoid – leads to very promising results:
In the case of the objective function having the form as in Theorem 3.28 (and many practical applications will fit into that scheme, since very often linear dependence on the perturbations is assumed) we get a special structure for the optimal solution set. If furthermore matrix A has full column rank n, i.e. each component of x is perturbed independently, we know that all optimal solutions are linearly dependent. Finally, the constraints describing the set X affect the set of solutions. For example a single constraint of the form $x^T\mathbf{1} = 1$ (a very common constraint in portfolio optimization problems in asset management) suffices to exclude linear multiples of an optimal solution and thus, $\mathcal{F}^*_{LRC}$ is a singleton which also implies continuity of the solution, hence stability of the problem in case a Slater point exists.

Remark 3.34. *A different approach – resulting in a similar objective as when robustifying using the robust counterpart method – dealing with ill-posed (i.e. not well-posed) problems is the* Tikhonov regularization*, see e.g. Kirsch [49]. There, the regularizing expression "$\alpha\|x\|^2$" is added to the objective of the original problem and the approximating program*

$$\begin{aligned} \min_{x\in X} \quad & f(x,u) + \alpha\|x\|^2 \\ s.t. \quad & g(x,u) \leq_K 0 \end{aligned} \tag{P_α}$$

is solved iteratively for $\alpha \to 0$. As the term $\|x\|^2$ is a strictly convex function, this guarantees uniqueness of the optimal solution of (P_α) and the sequence of solutions x^*_α for $\alpha \to 0$ converges to an optimal solution of problem (GCP_u).

3.4 Influence of the size of the uncertainty set

We have seen in the previous section that applying the robust counterpart with an ellipsoidal uncertainty set is promising with respect to achieving $\mathcal{U}$-stability of the program – assuming that some prerequisites are fulfilled. The second question, besides determining the shape of $\mathcal{U}$, is the question of the size and how much the robustification costs, meaning how much worse the optimal value f^* of the original problem becomes by using the robust counterpart formulation. That amount is likely to depend on the size of the uncertainty set and this interrelation is the subject of our investigations.

The main result in this subsection will be Theorem 3.37 which states explicitly how the size of the uncertainty set affects the optimal value $f^*_{LRC}(\hat{u}, \delta)$: The (asymptotic) costs of the robustification come at a linear rate in the size δ for $\delta \to 0$. This means that the increase in the optimal objective value when modifying $(GCP_{\hat{u}})$ to $(LRC_{\hat{u},\delta})$ is linear in δ. Before being able to prove that statement, we note the following intermediate results.

Lemma 3.35. *Let K be an ordering cone. Then there exists a point $c \in K$ such that $V_1(c) \subset K$.*

Proof. Let $z \in \operatorname{int} K$. There exists $\varepsilon > 0$ such that $V_\varepsilon(z) \subset K$. This is equivalent to

$$z + \tilde{w} \geq_K 0 \quad \forall \tilde{w}, \|\tilde{w}\| \leq \varepsilon$$

or, respectively, with $c := \frac{1}{\varepsilon} z$ and $w := \frac{1}{\varepsilon}\tilde{w}$

$$c + w \geq_K 0 \quad \forall w, \|w\| \leq 1$$

which is the desired result. □

Lemma 3.36. *Let $g(x, \cdot)$ be globally Lipschitz continuous with Lipschitz constant $L > 0$. Let furthermore $c \in K$ such that $V_1(c) \subset K$ and define*

$$\alpha := \alpha(\delta) = \delta\, L\, \operatorname{diam} \mathcal{U}'^{9}.$$

Then for each $x \in X$ satisfying

$$g(x, \hat{u}) + \alpha c \leq_K 0$$

[9]Recall that $\mathcal{U}'$ is defined as being equal to $\mathcal{U}$ but shifted such that $0 \in \mathcal{U}'$. Furthermore, let $\operatorname{diam} \mathcal{U}' = \max\limits_{v_1, v_2 \in \mathcal{U}'} \|v_1 - v_2\|$.

it also holds that

$$g(x,u) \leq_K 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}) = \hat{u} + \delta\mathcal{U}'.$$

Proof. Since $g(x,\cdot)$ is globally Lipschitz continuous with Lipschitz constant $L > 0$ it holds that

$$\|g(x,u) - g(x,\hat{u})\| \leq L\|u - \hat{u}\|.$$

With $u \in \mathcal{U}_\delta(\hat{u})$ we obtain the following chain of inequalities:

$$\|g(x,u) - g(x,\hat{u})\| \leq L\|u - \hat{u}\| \leq L\,\delta\,\operatorname{diam}\mathcal{U}' = \alpha \quad \forall u \in \mathcal{U}_\delta(\hat{u}). \tag{3.6}$$

The condition $g(x,\hat{u}) + \alpha c \leq_K 0$ is equivalent to $g(x,\hat{u}) \leq_K -\alpha c$. Furthermore, as $V_1(c) \subset K$,

$$c - \frac{1}{\alpha}w \geq_K 0 \quad \forall w \text{ with } \|w\| \leq \alpha.$$

This implies

$$g(x,\hat{u}) + w \leq_K -\alpha c + w = -\alpha(c - \frac{1}{\alpha}w) \leq_K 0 \quad \forall w \text{ with } \|w\| \leq \alpha.$$

Because of inequality (3.6) the point $w := g(x,u) - g(x,\hat{u})$ satisfies $\|w\| \leq \alpha$ for all $u \in \mathcal{U}_\delta(\hat{u})$ and thus

$$g(x,\hat{u}) + w = g(x,u) \leq_K 0 \quad \forall u \in \mathcal{U}_\delta(\hat{u}). \qquad \square$$

Using these preliminary results, we can now state and prove the theorem quantifying the costs associated with the advantage of obtaining a robust (and possibly unique and continuous) solution.

Theorem 3.37 (Costs of robustification). *Let f and g be globally Lipschitz continuous in u. Assume the existence of a Slater point for the program ($GCP_{\hat{u}}$), $\hat{u} \in \mathcal{U}$ and consider the corresponding local robust counterpart ($LRC_{\hat{u},\delta}$). Then there exists a $k > 0$ such that it holds for sufficiently small δ that*

$$f^*_{LRC}(\hat{u},\delta) \leq f^*(\hat{u}) + k\delta + o(\delta).$$

Proof. Because of Remark 3.4 we assume without loss of generality that the objective function is independent of u and linear in x, i.e. $f(x,u) = l(x)$. Note that when shifting the objective function into the set of constraints the properties of convexity and the existence of a Slater point are maintained. Furthermore, Lipschitz continuity of f and g also transfers to Lipschitz continuity of the new constraint function (Proposition F.1). Hence, we can without loss of generality prove the theorem for the local robust counterpart program in the following form:

$$\begin{aligned} \min_{x\in X} \quad & l(x) \\ \text{s.t.} \quad & g(x,u) \leq_K 0 \;\; \forall u \in \mathcal{U}_\delta(\hat{u}). \end{aligned} \tag{$LRC_{\hat{u},\delta}$}$$

We now introduce the auxiliary problem

$$\begin{aligned} \min_{x\in X} \quad & l(x) \\ \text{s.t.} \quad & g(x,\hat{u}) + \alpha c \leq_K 0 \end{aligned} \tag{P_{aux}}$$

with $\alpha := \delta\, L\, \operatorname{diam}\mathcal{U}' \geq 0$ (see Lemma 3.36) and $c \in K$ such that $V_1(c) \subset K$. Note that such a $c \in K$ exists because of Lemma 3.35. For δ – and thus α – sufficiently small, both programs $(LRC_{\hat{u},\delta})$ and (P_{aux}) have non-empty feasibility sets. Furthermore, Lemma 3.36 gives that each feasible point for (P_{aux}) is as well feasible for $(LRC_{\hat{u},\delta})$, i.e.

$$\mathcal{F}_{P_{aux}}(\alpha c) \subset \mathcal{F}_{LRC}(\hat{u},\delta)$$

with $\mathcal{F}_{P_{aux}}(\alpha c)$ denoting the feasible set of (P_{aux}) for given α and c, and thus

$$f^*_{LRC}(\hat{u},\delta) \leq f^*_{P_{aux}}(\alpha c) \tag{3.7}$$

where analogously $f^*_{P_{aux}}(\alpha c)$ denotes the optimal value of (P_{aux}). Note further that $f^*_{P_{aux}}(0) = f^*(\hat{u})$, since in the case $\alpha = 0$ the programs $(GCP_{\hat{u}})$ and (P_{aux}) coincide. As the existence of a Slater point for $(GCP_{\hat{u}})$ equivalently assures a Slater point for (P_{aux}) in the case of $\alpha = 0$, all the requirements for Proposition A.12 and Corollary A.13 are fulfilled and we thus obtain the following:

(i) The optimal value function of (P_{aux}), $f^*_{P_{aux}}(\alpha c)$, is Hadamard directionally differentiable at 0 for all directions d.

(ii) The value of the directional derivative at the point 0 is finite, i.e.

$$f^{*'}_{P_{aux}}(0;d) < \infty.$$

With the definition of the directional derivative (see Definition A.6), we get

$$f^{*'}_{P_{aux}}(0;d) = \lim_{t\downarrow 0} \frac{f^*_{P_{aux}}(0+td) - f^*_{P_{aux}}(0)}{t}$$

for any direction d, and using finiteness especially for the direction c we have

$$f^{*'}_{P_{aux}}(0;c) = \lim_{t\downarrow 0} \frac{f^*_{P_{aux}}(0+tc) - f^*_{P_{aux}}(0)}{t} =: k < \infty. \tag{3.8}$$

Furthermore, it holds that

- the directional derivative is positively homogeneous in d, see Lemma A.7, i.e.

$$f^{*'}_{P_{aux}}(0;\alpha c) = \alpha f^{*'}_{P_{aux}}(0;c) \overset{(3.8)}{=} \alpha k, \tag{3.9}$$

- as Hadamard directional differentiability implies Fréchet directional differentiability, see Proposition A.11, $f^*_{P_{aux}}(\alpha c)$ can be expressed as

$$f^*_{P_{aux}}(\alpha c) = f^*_{P_{aux}}(0) + f^{*'}_{P_{aux}}(0;\alpha c) + \underbrace{o(\|\alpha c\|)}_{=o(\alpha)}. \tag{3.10}$$

Combining all established results, we get

$$\begin{aligned} f^*_{LRC}(\hat{u},\delta) &\overset{(3.7)}{\leq} f^*_{P_{aux}}(\alpha c) \\ &\overset{(3.10)}{=} f^*_{P_{aux}}(0) + f^{*'}_{P_{aux}}(0;\alpha c) + o(\alpha) \\ &\overset{(3.9)}{=} f^*(\hat{u}) + \alpha k + o(\alpha) \end{aligned}$$

and finally, using the fact that α is linear in δ we end up with the desired result

$$f^*_{LRC}(\hat{u},\delta) \leq f^*(\hat{u}) + \tilde{k}\delta + o(\delta)$$

for sufficiently small δ. □

Thus, Theorem 3.37 allows to estimate and thus control the costs of a robustification using the robust counterpart approach. The inequality in Theorem 3.37 can be reformulated and thus implies the following for δ tending to zero:

$$\frac{f^*_{LRC}(\hat{u},\delta) - f^*(\hat{u})}{\delta} \to k.$$

This expression on the left hand side will be referred to as *relative performance gap* and is illustrated in the subsequent example.

Example 3.38. *Consider the following quadratic optimization problem in the n-dimensional case*

$$\min_{x\in X} \quad x^T\hat{\Sigma}x$$

with $X = \{x \in \mathbb{R}^n \mid x^T\mathbf{1} = 1\}$ *and* $\hat{\Sigma} \in \mathbb{R}^{n\times n}$ *symmetric and positive definite. The parameter* $\hat{\Sigma}$ *is not known exactly, but contains some uncertainty, e.g.* $\hat{\Sigma}$ *could represent a covariance matrix which was estimated from a finite data sample. As a simple local uncertainty set around the point* $\hat{\Sigma}$ *we choose the set containing all symmetric, positive semidefinite matrices* Σ *(recall that positive semidefiniteness will be denoted by* $\Sigma \succeq 0$*) that are "lying close" to* $\hat{\Sigma}$*, where closeness is measured using the trace norm. Thus, the set is described by*

$$\mathcal{U}_\delta(\hat{\Sigma}) = \left\{\Sigma \in \mathbb{R}^{n\times n} \mid \|\Sigma - \hat{\Sigma}\|_{tr} \leq \delta, \Sigma = \Sigma^T, \Sigma \succeq 0\right\}.$$

To determine the explicit reformulation of the robust counterpart program, we need to find the worst case parameter Σ_{wc}, i.e. the solution of

$$\max_{\Sigma \in \mathcal{U}_\delta(\hat{\Sigma})} \quad x^T \Sigma x$$

or equivalently

$$\begin{aligned} \max_{\substack{\Sigma \in S^n \\ \Sigma \succeq 0}} \quad & x^T \Sigma x \\ s.t. \quad & \|\Sigma - \hat{\Sigma}\|_{tr}^2 \leq \delta^2. \end{aligned}$$

Using the notation $\langle A, B \rangle = tr(AB)$ (see also Remark 2.7) to denote the inner product of the space of symmetric $n \times n$ matrices, S^n, the problem can be reformulated as

$$\begin{aligned} \max_{\substack{\Sigma \in S^n \\ \Sigma \succeq 0}} \quad & \langle \Sigma, xx^T \rangle \\ s.t. \quad & \langle \Sigma - \hat{\Sigma}, \Sigma - \hat{\Sigma} \rangle \leq \delta^2. \end{aligned}$$

Introducing $C := \Sigma - \hat{\Sigma}$, i.e. $\Sigma = \hat{\Sigma} + C$, we furthermore obtain

$$\begin{aligned} \max_{\substack{C \in S^n \\ C + \hat{\Sigma} \succeq 0}} \quad & \langle \hat{\Sigma}, xx^T \rangle + \langle C, xx^T \rangle \\ s.t. \quad & \langle C, C \rangle \leq \delta^2. \end{aligned}$$

As the expression $\langle \hat{\Sigma}, xx^T \rangle$ is independent of the variable C, it suffices to maximize $\langle C, xx^T \rangle$. We furthermore solve a relaxed optimization problem by neglecting the constraint $C + \hat{\Sigma} \succeq 0$ for a moment. The inner product $\langle C, xx^T \rangle$ is maximized if the arguments are multiples, hence the optimal solution is given by

$$C^* = \delta \frac{xx^T}{\|xx^T\|_{tr}} = \delta \frac{xx^T}{\|x\|_2^2}$$

by using that

$$\|xx^T\|_{tr} = \sqrt{tr(xx^T xx^T)} = \sqrt{tr(x^T xx^T x)} = \sqrt{(x^T x)^2} = \|x\|^2.$$

Thus, the worst case matrix Σ_{wc} is finally given by

$$\Sigma_{wc} = \hat{\Sigma} + \delta \frac{xx^T}{\|x\|^2}$$

which is a symmetric and positive semidefinite matrix, i.e. the previously neglected condition is fulfilled as well. With this result the robust counterpart program can

be reformulated to

$$\begin{aligned}
&\min_{x\in X} \max_{\Sigma\in\mathcal{U}_\delta(\hat{\Sigma})} \quad x^T\Sigma x \\
&= \min_{x\in X} \quad x^T\left(\hat{\Sigma} + \delta\frac{xx^T}{\|x\|^2}\right)x \\
&= \min_{x\in X} \quad x^T\hat{\Sigma}x + \frac{\delta}{\|x\|^2}\underbrace{x^Txx^Tx}_{=(\|x\|^2)^2} \\
&= \min_{x\in X} \quad x^T\hat{\Sigma}x + \delta\|x\|^2 \\
&= \min_{x\in X} \quad x^T(\hat{\Sigma} + \delta I)x.
\end{aligned}$$

with I the $n \times n$ identity matrix.

Solving both the original and the robust program for various values of δ yields the illustration of the relative performance gap as shown in Figure 3.11. The value of $k = \lim_{\delta\to 0} \frac{f^*_{LRC}(\hat{\Sigma},\delta) - f^*(\hat{\Sigma})}{\delta}$ is approximately 0.75 in this example.

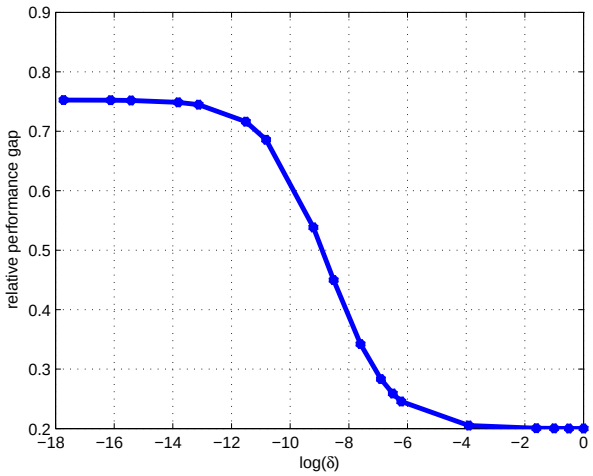


Figure 3.11: Illustration of the relative performance gap.

Remark 3.39. *Using the $o(\delta)$ notation, the inequality in Theorem 3.37 implies*

$$\lim_{\delta\to 0}\left|\frac{f^*_{LRC}(\hat{u},\delta) - f^*(\hat{u}) - k\delta}{\delta}\right| \leq \lim_{\delta\to 0}\frac{o(\delta)}{\delta} = 0$$

or equivalently

$$\left|\frac{f^*_{LRC}(\hat{u},\delta) - f^*(\hat{u}) - k\delta}{\delta}\right| \leq l \quad \forall\delta \in V_\varepsilon(0),$$

i.e. for δ sufficiently small. This gives

$$\|f^*_{LRC}(\hat{u}, \delta) - f^*(\hat{u})\| \leq (k + l)\delta \quad \forall \delta \in V_\varepsilon(0).$$

Thus Theorem 3.37 implies pointwise Lipschitz continuity of the extreme value function in $\delta = 0$.

The above result of Theorem 3.37 applies to our general convex optimization problem, but it has one rather strong requirement: the existence of a Slater point. Ben-Tal and Nemirovksi [4] proved the same result for linear problems. Their result is derived without the assumption of a Slater point, but under (similar) technical constraints.

At the end of this chapter we want to shortly summarize the main results. First of all, we have proved that all the stability properties are maintained when the original problem (GCP_u) is changed to the (local) robust counterpart program. We then have found (see Theorem 3.28 and Corollary 3.31) that using an ellipsoidal uncertainty set in the robust counterpart approach reduces the set of optimal solutions to (mainly) linearly dependent ones. Having further constraints in the optimization problem often leads to a single optimal solution in practical applications which is then also continuous according to Theorem 3.21, part (iv). Thus, applying the robust counterpart with an ellipsoidal uncertainty set yields the desired properties of a unique and continuous optimal solution. Finally, we have seen (see Theorem 3.37) that for small δ the increase in the optimal objective value compared to the non-robust solution – i.e. the costs of robustification – is linear in the size δ. Hence, using a rather small uncertainty set for robustification already gives the benefit of a unique and continuous solution while still keeping the costs controllable.

Part II

Application of robust optimization in asset management

Chapter 4

Traditional portfolio optimization

4.1 Introduction

In the second part of this dissertation we want to apply the robust counterpart approach to a famous optimization problem in finance: the Markowitz portfolio optimization problem, introduced by Markowitz in 1952, see [56]. Before eventually formulating and solving the optimization problem, we describe the underlying financial market and illustrate the necessity of robustification.

We consider a financial market that consists of n risky assets. In our examples the market is given by the five assets Lehman Euro Aggregate, DJ Stoxx 50, DJ Stoxx Small Caps Europe, MSCI Japan and MSCI Emerging Markets. The first index is a bond index whereas the other four are stock indices which usually are a lot riskier (i.e. much more volatile) than bonds. This setting represent an asset universe that can e.g. be used for strategical asset allocation where different markets are assumed to be represented by selected indices capturing the main characteristics of the respective economy. To get an idea about the individual assets, Figure 4.1 illustrates the performance of the different asset classes over the time period from July 2001 to December 2005, where for easier comparability all assets were scaled to start at a value of 1 at the beginning of the underlying time period.

The difference in the general behavior of stocks and bonds can nicely be seen in the historical performance since the Lehman Euro resembles a slightly upward sloping line in contrast to the rather heavily moving stocks. As can also be observed, at the beginning of the considered time period there was a bear market, resulting in a negative performance of the stocks. Later on, the situation changes to a bull market with quite attractive gains in stock investments. The figure furthermore shows that the stock indices tend to behave similarly which indicates a (high) positive correlation.

The characteristics of the market are described in terms of the expected returns of the individiual assets, their risk measured by the standard deviation

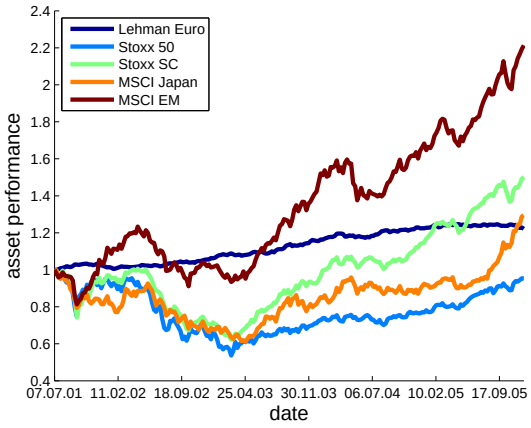


Figure 4.1: Historical performance of the asset classes (07/2001 - 12/2005).

(volatility), and finally the correlation between the assets. The assets' volatilities and the correlation are captured together in the covariance matrix. The vector of expected returns (denoted by μ) and the covariance matrix (denoted by Σ) are thus the parameters that eventually enter the portfolio optimization problem. To obtain particular parameter values for the optimization, *estimators* need to be calculated. Mostly the maximum likelihood estimators (MLEs) based on a historical sample are used. In the following, we illustrate that the MLEs can attain a rather wide spectrum of different estimated values when calculated based on changing data samples.

Having the historical data sample as shown above, we move through time and always use a year's time (i.e. the directly preceding 52 data points) to calculate the maximum likelihood estimators for the individual expected asset returns and volatilites. To point out the extremes, Tables 4.1 and 4.2 contain the annualized[1] mean and volatility and the correlation matrix for selected time periods during the bear market and the bull market, represented by the weeks from 13.04.02 to 05.04.03 and from 03.07.04 to 25.06.05, respectively.

The asset characteristics in the different market phases are additionally illustrated in Figure 4.2, where we have plottet the expected return and the standard deviation (volatility) of the individual assets in a risk-return diagram. It can be seen that the bond index, the Lehman Euro, remains rather unaffected by the general market situation, since its expected return and its volatility are quite the same in both a bear and a bull market. In the time period representing the bear market, the stock indices are clustered in the lower right corner, expressing that they have large negative expected returns and the tendency to higher risk. The

[1]Annualization has been done by multiplying average weekly returns by 52 and standard deviation by $\sqrt{52}$. This simply scales the weekly setting to an annual point of view which is used here only for reporting.

	return	volatility	correlation				
Lehman Eur	6.5%	2.4%	1.00	-0.45	-0.18	-0.03	-0.06
Stoxx 50	-40.7%	17.6%	-0.45	1.00	0.74	0.34	0.44
Stoxx SC	-41.7%	16.6%	-0.18	0.74	1.00	0.54	0.74
MSCI Japan	-31.5%	20.1%	-0.03	0.34	0.54	1.00	0.52
MSCI EM	-21.4%	16.7%	-0.06	0.44	0.74	0.52	1.00

Table 4.1: Annualized returns and volatilities and the correlation in a bear market.

	return	volatility	correlation				
Lehman Eur	5.5%	2.7%	1.00	-0.25	-0.16	-0.09	-0.13
Stoxx 50	14.8%	9.5%	-0.25	1.00	0.72	0.58	0.60
Stoxx SC	22.5%	8.6%	-0.16	0.72	1.00	0.56	0.57
MSCI Japan	1.4%	11.5%	-0.09	0.58	0.56	1.00	0.72
MSCI EM	22.4%	10.9%	-0.13	0.60	0.57	0.72	1.00

Table 4.2: Annualized returns and volatilities and the correlation in a bull market.

upper left side of the diagram shows the high returns which can be expected in a bull market.

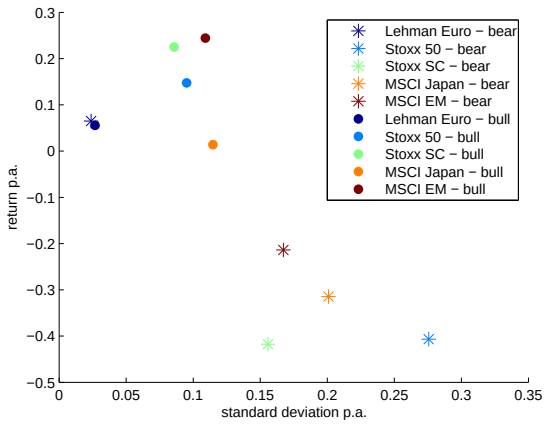
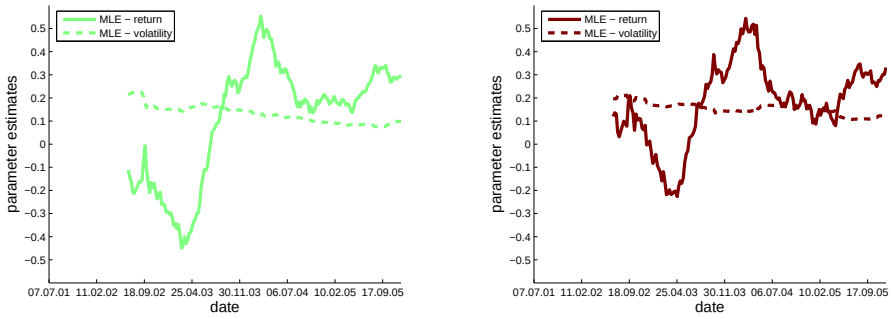


Figure 4.2: Asset characterization in bear and bull markets.

Tables 4.1 and 4.2 give an indication of the extreme parameter values that can be obtained, but we are also interested in the changes in the estimators over time. Figures 4.3 plots the maximum likelihood estimators for the return and the volatility in case of the stock indices Stoxx SC (Figure 4.3(a)) and MSCI Emerging Markets (Figure 4.3(b)). The other two stock indices showed similar plots for their

estimators. Note that compared to the plot of the historical asset performance, the estimators can only be calculated after the first year of data. It can nicely be seen that the MLEs for the return undergo rather drastic changes over time, a fact that suggests that the return vector is a rather uncertain parameter in the optimization problem later on. The volatility estimators of the stock indices also change over time, but not as heavily as the return estimates.



(a) Stoxx SC (b) MSCI EM

Figure 4.3: Estimators for return and volatility of stock indices over time.

The historical performance in Figure 4.1 already indicated that the bond index is a lot less riskier than the stocks, a fact that is also reflected in the stability of the respective parameter estimates, shown in Figure 4.4. For better comparison the axes are scaled analogous to the ones in Figure 4.3.

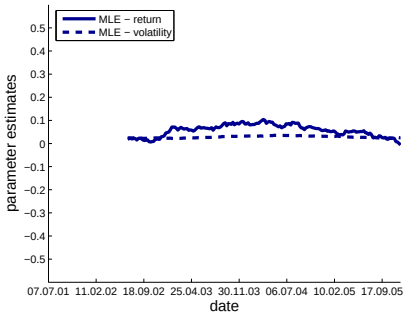
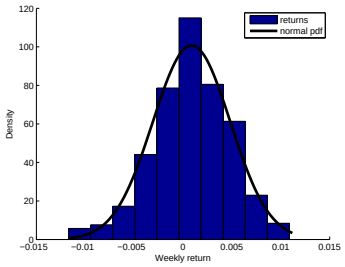


Figure 4.4: Estimators for return and volatility of the bond index over time.

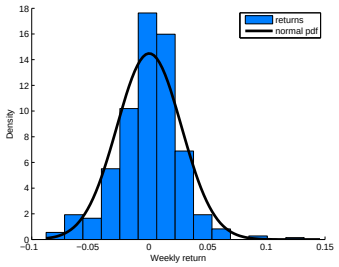
To model a financial market, we assume that the vector of asset returns follows a (multivariate) distribution. In many practical applications it is simply assumed that the asset returns are normally distributed. This might be a good approximation for some asset classes, e.g. for bond indices, but for for other types of assets, normality of the respective return vectors is often violated and hence, modeling a market by a normal distribution is not always sufficient, e.g. if fat

tails need to be explicitly considered. The class of elliptically symmetric distributions represents a more general framework and thus allows for more flexibility in modeling returns. Section 4.2 below introduces elliptical distributions and their properties.

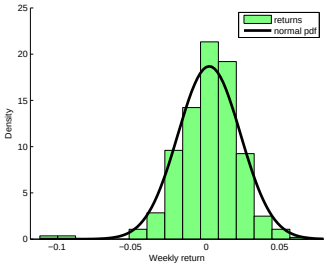
Before, we want to test whether our particular data sample fits into the general assumption that it follows an elliptical distribution. We consider the sample of each asset and test the hypothesis that the data come from a normal distribution. Using a χ^2-goodness-of-fit test (see e.g. [80]), the hypothesis that the individual data samples stem from a normal distribution can only be rejected for the Stoxx SC index at the 5% level of significance.



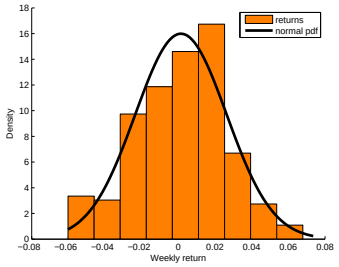
(a) Lehman Euro



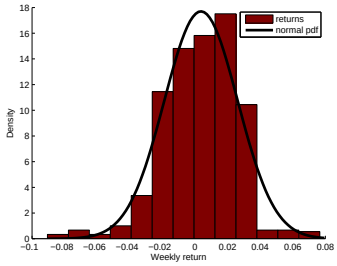
(b) Stoxx 50



(c) Stoxx SC



(d) MSCI Japan



(e) MSCI EM

Figure 4.5: Histograms of the asset returns.

Figure 4.5 shows histograms of the weekly returns together with the probability density function fitted into the data. The histogram for the Stoxx SC index (Figure 4.5(c)) seems rather normally distributed at first sight, but the outliers to

the left probably led to the rejection of the hypothesis. More sophisticated elliptical distribution (e.g. the normal inverse Gaussian (NIG) distribution) are then needed to model such a behavior. Hence, summarizing it can be said that the assumption of the returns following an elliptical distribution seems appropriate for our data sample.

The Markowitz portfolio optimization problem is based on a mean-variance setting which only used the mean and the variance to characterize (and optimize) a portfolio. Thus, after having a distributional model for the financial market, we need to define estimators for the first two moments of the assumed distribution which are then used as input parameters in the portfolio optimization problem. Besides the widely used maximum likelihood estimators, there exist other statistical estimators as well, especially for the mean of an elliptical distribution. As the estimates themselves can vary quite a bit depending on the data sample (see Figure 4.3), it might be interesting to use more robust parameter estimators which e.g. omit outliers in the sample. We will present selected estimators especially for the mean vector of an elliptical distribution in Section 4.3.

Section 4.4 reviews portfolio theory and defines efficient portfolios and the associated efficient frontier based on the underlying financial market. The efficient portfolios are obtained by solving a parametric optimization problem with the input parameters being the vector of expected asset returns and the covariance matrix, hence the market characteristics. This (classical) portfolio optimization problem will be shown to be convex and thus the theoretical results from the first part of the dissertation are applicable (in the special case of the cone being the standard cone). Since it can furthermore be proved that the optimal solution is unique, Theorem 2.45 gives continuity of the solution of the classical portfolio problem with respect to the parameters, a result that will be needed again in Chapter 6. Despite having continuity, the optimal portfolio crucially depends on the input parameters, see e.g. Jorion [48] or Best and Grauer [13], and especially on the assumed performance of the assets (i.e. the vector of expected asset returns), see e.g. Chopra and Ziemba [21]. This effect can best be seen in case of the maximum return portfolio, the portfolio which completely ignores the associated risk and relies solely on the estimated asset performance.

Hence, having seen in Figure 4.3 that the parameter estimates (especially the expected return) change heavily through time, and knowing that the optimization result strongly depends on these estimates, it seems both necessary and natural to seek for more stable solutions. In Chapter 5 we will apply the robust counterpart approach to the classical portfolio optimization problem and investigate the achievements obtained thereof. The main problem in practical applications is the definition of appropriate uncertainty sets for solving the robust formulation. We present and discuss two different approaches of creating ellipsoidal[2]

[2]Recall that we have shown in Theorem 3.28 that ellipsoidal uncertainty sets seem to be more promising.

uncertainty sets. The first approach is to consider a confidence ellipsoid for the parameter estimates which is based on the distribution of the considered point estimates, e.g. the maximum likelihood estimators. The second idea is to use several statistical estimator to create an uncertainty set.

4.2 Elliptical distributions

In this section we introduce some fundamental characteristics of elliptical distributions that can be used to model a financial market. We only collect the basic definitions and properties, and refer to Fang, Kotz and Ng [29] and Fang and Zhang [30] for further results and more details.

Definition 4.1. *A random vector $R \in \mathbb{R}^n$ is said to have a* spherical *distribution if*

$$OR \stackrel{d}{=} R$$

for every orthogonal matrix $O \in \mathbb{R}^{n\times n}$, and with "$\stackrel{d}{=}$" denoting equality of distributions.

The following theorem summarizes some useful equivalence properties for the basic class of spherical distributions which will later be extended to the class of elliptical distributions.

Theorem 4.2. *Let $R \in \mathbb{R}^n$ be a random vector. Then, the following statements are equivalent:*

(i) $OR \stackrel{d}{=} R$ *for every orthogonal matrix $O \in \mathbb{R}^{n\times n}$.*

(ii) There exists a function $\phi : \mathbb{R} \to \mathbb{R}$, called the characteristic generator*, such that the characteristic function ψ of R has the form*

$$\psi(t) = \mathbf{E}\left[e^{itR}\right] = \phi(t^T t).$$

(iii) The vector R has a stochastic representation of the form

$$R \stackrel{d}{=} Zu^{(n)}$$

with the generating random variable $Z \in \mathbb{R}, Z \geq 0$ being independent of $u^{(n)}$, a uniformly distributed random vector on the unit sphere in $\mathbb{R}^n$.

Proof. See Fang, Kotz and Ng [29], Theorem 2.5. □

A spherically distributed random vector R does not necessarily have a probability density function (pdf). But in case a density function $\varphi_R : \mathbb{R}^n \to \mathbb{R}$ exists, it must be of the form $\varphi_R(x) = \xi_R(x^T x)$ (analogous to $\phi(t^T t)$) for some $\xi_R : \mathbb{R}_+ \to \mathbb{R}_+$ which is called the *density generator*[3]. Furthermore, we obtain the following results.

Proposition 4.3. *Let $R \stackrel{d}{=} Zu^{(n)}$ be spherically distributed. Then, R has a density generator $\xi_R : \mathbb{R} \to \mathbb{R}$ if and only if the generating variate Z has a probability density function $\varphi_Z : \mathbb{R} \to \mathbb{R}$. Furthermore, the relationship between these two functions is analytically given by*

$$\varphi_Z(z) = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \cdot z^{n-1} \xi_R(z^2).$$

Additionally, if R possesses a probability density function, then all the marginal densities exist as well.

Proof. See Fang, Kotz and Ng [29], Theorems 2.9 and 2.10. □

Remark 4.4. *Inverting the formula in the above proposition, we can equivalently express the density generator of R in terms of the pdf of Z by*

$$\xi_R(t) = \frac{\Gamma(\frac{n}{2})}{2\pi^{\frac{n}{2}}} \cdot t^{-\frac{n-1}{2}} \varphi_Z(\sqrt{t}).$$

Notation 4.5. *To denote that the vector $R \in \mathbb{R}^n$ is spherically distributed with the characteristic generator ϕ, we will write $R \sim \mathcal{S}_n(\phi)$. When dealing with a density generator ξ, this will analogously be denoted by $R \sim \mathcal{S}_n(\xi)$.*

After having briefly introduced spherical distributions, we now extend the concept to elliptically symmetric distributions. In the literature elliptically symmetric distributions are often called "elliptically contoured distributions", as the level curves of the density (e.g. in a contour plot) are ellipses. In the following we will simply use the term *elliptical distributions* instead of elliptically symmetric distributions or elliptically contoured distributions.

Definition 4.6. *A random vector $R \in \mathbb{R}^n$ is said to be* elliptically *distributed with the parameters $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ if*

$$R \stackrel{d}{=} \mu + A^T Y, \quad Y \sim S_k(\phi)$$

with $A \in \mathbb{R}^{k \times n}$ such that $A^T A = \Sigma$ and rank$(\Sigma) = k$. To abbreviate R being elliptically distributed with the characteristic generator ϕ, we will write $R \sim \mathcal{E}_n(\mu, \Sigma, \phi)$.

[3] Note that both in the book of Fang and Zhang [30] and in Fang, Kotz and Ng [29] the letter to denote the probability density function and the density generator is the same. The generator has the (scalar) argument $x^T x$, and the pdf has the argument x, the function description otherwise is the same – as can be seen from the equation $\varphi_R(x) = \xi_R(x^T x)$.

Remark 4.7. *Note that the spherical distribution equals the elliptical distribution with $\mu = 0$ and $A = \Sigma = I$.*

Similar to the above Theorem 4.2 we get the following statements with respect to elliptical distributions.

Theorem 4.8. *Let $R \sim \mathcal{E}_n(\mu, \Sigma, \phi)$ and let rank$(\Sigma) = k$. It holds:*

(i) There exists a function $\phi : \mathbb{R} \to \mathbb{R}$ such that the characteristic function ψ of R has the form

$$\psi(t) = \mathbf{E}\left[e^{it^T R}\right] = e^{it^T \mu}\phi(t^T \Sigma t).$$

(ii) The vector R has a stochastic representation of the form

$$R \stackrel{d}{=} \mu + Z A^T u^{(k)}$$

with $Z \geq 0$ being independent of $u^{(k)}$ and $A^T A = \Sigma$.

Proof. See Fang, Kotz and Ng [29], page 32. □

Remark 4.9. *Any scalar function ϕ fulfilling a certain integrability condition (for the exact condition, see [29] or [30]) can determine an elliptical distribution ([30], Theorem 2.6.1). As ϕ is therefore not unique, we can without loss of generality assume that ϕ is chosen such that*

$$-2\phi'(0) = 1 \tag{4.1}$$

holds, see Fang and Zhang [30], page 67.

The next proposition summarizes several useful results about the moments, marginals and combinations of elliptical distributions.

Proposition 4.10. *Let $R \sim \mathcal{E}_n(\mu, \Sigma, \phi)$ and $\mathbf{E}[Z^2] < \infty$ with Z as given in the representation formula in Theorem 4.8, part (ii). Then, the following holds:*

(i) The expected value and the covariance matrix of R are given by

$$\begin{aligned} \mathbf{E}[R] &= \mu, \\ \mathbf{Cov}[R] &= \frac{\mathbf{E}[Z^2]}{rank(\Sigma)} \cdot \Sigma = -2\phi'(0)\Sigma = \Sigma. \end{aligned}$$

where the last equality holds due to the normalization assumption in Equation 4.1 in the above remark.

(ii) Any linear transformation of an elliptically distributed variable is again elliptically distributed, more precisely:
Let $R \sim \mathcal{E}_n(\mu, \Sigma, \phi)$, rank$(\Sigma) = k$, $B \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then

$$BR + b \sim \mathcal{E}_m(B\mu + b, B\Sigma B^T, \phi).$$

(iii) Any marginal distributions of an elliptically distributed variable are again elliptical, more precisely: Let $R \sim \mathcal{E}_n(\mu, \Sigma, \phi)$ and partition R, μ and Σ into

$$R = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

with appropriate dimensions k and $n-k$ such that $R_1 \in \mathbb{R}^k$ and $R_2 \in \mathbb{R}^{n-k}$. Then it holds that

$$\begin{aligned} R_1 &\sim \mathcal{E}_k(\mu_1, \Sigma_{11}, \phi), \\ R_2 &\sim \mathcal{E}_{n-k}(\mu_2, \Sigma_{22}, \phi). \end{aligned}$$

(iv) The conditional distribution of an elliptically distributed variable is again elliptical. Formally, this is stated as follows:
Let $R \stackrel{d}{=} \mu + ZA^T u^{(n)} \sim \mathcal{E}_n(\mu, \Sigma, \phi)$ with $\Sigma = A^T A$ being positive definite. Consider again the partitioning as given in part (iii). Then it holds that

$$(R_1 | R_2 = x_2) \sim \mathcal{E}_k(\tilde{\mu}_1, \tilde{\Sigma}_1, \tilde{\phi})$$

with

$$\tilde{\mu}_1 = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \tag{4.2}$$

$$\tilde{\Sigma}_1 = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \tag{4.3}$$

and $\tilde{\phi}$ appropriate (for details see [29], page 45).

(v) Let $R_s \sim \mathcal{E}_n(\mu, \Sigma, \phi), s = 1, \ldots, S$ independent and identically distributed. Then it holds that

$$Y = \sum_{s=1}^{S} R_s \sim \mathcal{E}_n(S\mu, \Sigma, \phi^S)$$

with $\phi^S = \prod_{s=1}^{S} \phi$.

Proof. See Fang, Kotz and Ng [29], Section 2.5 for the parts (i) to (iv), part (v) follows from Theorem 4.1 in Hult and Lindskog [39]. □

The moments of an elliptical distribution are needed in our application of portfolio optimization when determining parameter estimates for the vector of expected returns and the covariance matrix, the input parameters of the optimization problem. Furthermore, the marginals being again elliptical guarantees the proper modelling of the individual assets. Finally, explicitly having the distribution of a sum of independent and identically elliptically distributed variables allows us to describe the distribution of selected parameter estimates. For example, with the formula of part (v) it is known that the maximum likelihood estimator for the mean follows again an elliptical distribution if the realizations in the sample of historical data are elliptically distributed, and furthermore, the moments are given as well. Hence, we can use this information to create a confidence ellipsoid and use this as an uncertainty set for the vector of expected returns. This will be done in more detail in Section 5.2.

Remark 4.11. *As in the case of spherical distributions, an elliptically distributed variable does not necessarily have a probability density function. If a density exists, then it must hold that rank*$(\Sigma) = n$*. Furthermore, as the probability density function of* $Y \sim \mathcal{S}_n(\phi)$ *is of the form* $\varphi_Y(y) = \xi_Y(y^T y)$*, the pdf of* $R = \mu + A^T Y \sim \mathcal{E}_n(\mu, \Sigma, \phi)$ *is of the form*

$$\varphi_R(x) = |\Sigma|^{-\frac{1}{2}} \xi_Y\left((x-\mu)^T \Sigma^{-1}(x-\mu)\right),$$

see Fang, Kotz and Ng [29], page 46.

A sometimes useful result gives the following proposition which links the density function of the elliptically distributed random variable and the density of its generating variate, similar to Proposition 4.3.

Proposition 4.12. *Let* $R \sim \mathcal{E}_n(\mu, \Sigma, \phi)$ *with* $\Sigma = A^T A$ *positive definite, and let* R *possess a density function. Then* R *can be represented as* $R \overset{d}{=} \mu + Z A^T u^{(n)}$ *(Theorem 4.8). Assume furthermore that the cumulative density function (cdf) of* Z *is absolutely continuous (hence,* Z *possesses a probability density function). Then, the probability density function* φ_R *of* R *is given by*

$$\varphi_R(x) = \sqrt{\det(\Sigma^{-1})} \cdot \xi_Z\left((x-\mu)^T \Sigma^{-1}(x-\mu)\right), \quad x \neq \mu$$

with

$$\xi_Z(t) := \frac{\Gamma\left(\frac{n}{2}\right)}{2\pi^{\frac{n}{2}}} \cdot t^{-\frac{n-1}{2}} \cdot \varphi_Z(\sqrt{t}).$$

Proof. See Frahm [31], Corollary 4. □

Using this just stated result about the explicit expression of the density function, it is rather straightforward to show symmetry with respect to the mean μ. This fact is of importance in the subsequent sections, as we will be investigating different estimators for μ which are only meaningful substitutes for the mean in case of symmetric distributions.

Proposition 4.13. *Let $R \sim \mathcal{E}_n(\mu, \Sigma, \phi)$ with $\Sigma = A^T A$ positive definite, let R possess a density function and let Z (the generating variate of R) possess a density function. Then the probability density function of R is symmetric with respect to the mean vector $\mu = \mathbf{E}[R]$.*

Proof. From Proposition 4.12 we have that the probability density function can be expressed as

$$\varphi_R(x) = \sqrt{\det(\Sigma^{-1})} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{2\pi^{\frac{n}{2}}} \cdot \left((x-\mu)^T\Sigma^{-1}(x-\mu)\right)^{-\frac{n-1}{2}} \cdot \varphi_Z(\sqrt{(x-\mu)^T\Sigma^{-1}(x-\mu)})$$

with φ_Z being the density of the generating variate Z. As it holds for all $x \in \mathbb{R}^n, x \neq 0$ that

$$\begin{aligned}\varphi_R(\mu - x) &= \varphi_R(\mu + x) \\ &= \sqrt{\det(\Sigma^{-1})} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{2\pi^{\frac{n}{2}}} \cdot \left(x^T\Sigma^{-1}x\right)^{-\frac{n-1}{2}} \cdot \varphi_Z(\sqrt{x^T\Sigma^{-1}x}),\end{aligned}$$

symmetry with respect to μ is proved. □

To close the section about elliptical distribution, we use the multivariate standard normal distribution to explicitly state the various generators and all the different introduced notations and calculations linking them.

Example 4.14. *We start with a multivariate standard normally distributed random variable. Let $Y \sim \mathcal{N}(0, I)$, i.e. $Y \sim \mathcal{S}_n(\phi)$ for some characteristic generator ϕ. From Theorem 4.2 we have that Y can be expressed as $Y = Zu^{(n)}$. Furthermore, as the normal distribution is a continuous distribution, it holds that $\mathbf{P}(Y = 0) = 0$. With these two prerequisites Corollary 1 on page 57 in Fang and Zhang [30] states that*

$$Z \overset{d}{=} \|Y\|.$$

As $W := \|Y\|^2 = Y^T Y \in \mathbb{R}$ is known to follow a χ^2_n-distribution with the pdf (see e.g. [46], page 416)

$$\varphi_W(w) = \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} e^{-\frac{w}{2}} w^{\frac{n}{2}-1},$$

the probability density function of Z is obtained by a transformation of the density and calculates to

$$\varphi_Z(z) = \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} \cdot 2e^{-\frac{z^2}{2}} z^{n-1} \quad \forall z \in \mathbb{R}_+.$$

Using the equation in Remark 4.4 which links the density generator ξ_Y of Y to the pdf φ_Z, ξ_Y is given by:

$$\begin{aligned}\xi_Y(t) &= \frac{\Gamma(\frac{n}{2})}{2\pi^{\frac{n}{2}}} t^{-\frac{n-1}{2}} \cdot \varphi_Z(\sqrt{t}) \\ &= \frac{\Gamma(\frac{n}{2})}{2\pi^{\frac{n}{2}}} t^{-\frac{n-1}{2}} \left[\frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} \cdot 2e^{-\frac{t}{2}} t^{\frac{n-1}{2}} \right] \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{t}{2}}\end{aligned}$$

using $t := z^2$. Thus finally, the density of Y is given by

$$\varphi_Y(y) = \xi_Y(y^T y) = \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{y^T y}{2}},$$

the well-known formula for the density of a standard normally distributed variable.

Additionally, we can determine the characteristic generator from the characteristic function of $Y \sim \mathcal{N}(0, I)$. The characteristic function of Y is given by (see e.g. Fang and Zhang, Theorem 2.3.1)

$$\psi_Y(y) = e^{-\frac{1}{2} y^T y}$$

and by recalling the relation $\psi_Y(y) = \phi(y^T y)$, the characteristic generator ϕ is

$$\phi(t) = e^{-\frac{t}{2}}.$$

Note that in this case of a standard normal distribution, the assumption $-2\phi'(0) = 1$ is fulfilled, as $\phi'(0) = -\frac{1}{2}$.

Finally, we consider a multivariate normally distributed random variable $R = \mu + A^T Y \sim \mathcal{N}(\mu, \Sigma)$ with $\Sigma = A^T A$ a positive definite matrix. From Remark 4.11 we straightforwardly obtain the density function φ_R for R from the density generator ξ_Y for Y by

$$\begin{aligned}\varphi_R(x) &= |\Sigma|^{-\frac{1}{2}} \xi_Y\left((x-\mu)^T \Sigma^{-1} (x-\mu)\right) \\ &= |\Sigma|^{-\frac{1}{2}} \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}\left((x-\mu)^T \Sigma^{-1} (x-\mu)\right)}.\end{aligned}$$

This matches the formula in Definition D.1.

4.3 Parameter estimation

As already mentioned, parameter estimates representing the expectation and the covariance matrix of the financial asset returns are needed as input for the portfolio optimization problem. Since we assume that the asset returns are modeled according to an elliptical distribution, we need estimators for the first two moments

of an elliptically distributed random vector. The most widely used estimator for the mean in practical applications is the maximum likelihood estimator(MLE) or sample estimator. Other estimators are e.g. proposed by Jorion [47] who uses a Bayesian estimators instead of the MLE, or by Jobson and Korkie [45] who suggest Stein-type estimator for obtaining more stable results. More robust estimators are e.g. introduced in Perret-Gentil and Victoria-Feser [64].

In this section we present and investigate selected statistical estimators for the mean of elliptical distributions. As estimator for the covariance matrix we will merely consider the maximum likelihood estimator. Furthermore, some main characteristics of these estimators are summarized.

To be able to define the estimators properly, let a sample of length S of i.i.d. random vectors following an elliptical distribution be given. This sample will be denoted by $R_1, \dots, R_S$ with $R_s \in \mathbb{R}^n$, $R_s \sim \mathcal{E}_n(\mu, \Sigma, \phi)$, $s = 1, \dots, S$, and the characteristic generator ϕ being chosen such that $-2\phi'(0) = 1$ holds, see Remark 4.9. We will furthermore assume that for the elliptical distributions under consideration a density function exists, thus we can equivalently write $R_s \sim \mathcal{E}_n(\mu, \Sigma, \xi)$, $s = 1, \dots, S$ with ξ being the according density generator.

The most widely used estimators for the mean μ and the covariance matrix Σ are the sample estimators or the maximum likelihood estimators:

Definition 4.15. *Let $R_s \sim \mathcal{E}_n(\mu, \Sigma, \phi)$, $s = 1, \dots, S$ i.i.d. The* maximum likelihood estimators *for μ and Σ based on the sample of length S are given by*

$$\hat{\mu}_S^{ML} := \frac{1}{S} \sum_{s=1}^{S} R_s,$$
$$\hat{\Sigma}_S^{ML} := \frac{1}{S} \sum_{s=1}^{S} (R_s - \hat{\mu}_S^{ML})(R_s - \hat{\mu}_S^{ML})^T.$$

As the sample estimator for the mean coincides with $\hat{\mu}_S^{ML}$, and the sample covariance is a constant multiple of $\hat{\Sigma}_S^{ML}$ (with the constant $\frac{S}{S-1}$), we will not pursue the explicit investigation of the sample estimators any further.

In the particular setting that the random vectors of the sample are normally distributed, the (joint) distribution of the maximum likelihood estimators $\hat{\mu}_S^{ML}$ and $\hat{\Sigma}_S^{ML}$ is analytically given as described in the following proposition.

Proposition 4.16. *Let $R_s \sim \mathcal{N}(\mu, \Sigma)$, $s = 1, \dots, S$ i.i.d. Then, the maximum likelihood estimators $\hat{\mu}_S^{ML}$ and $\hat{\Sigma}_S^{ML}$ are independent and distributed according to*

$$\hat{\mu}_S^{ML} \sim \mathcal{N}\left(\mu, \frac{1}{S}\Sigma\right), \quad \hat{\Sigma}_S^{ML} \sim \mathcal{W}\left(\frac{1}{S}\Sigma, S-1\right)$$

with $\mathcal{W}(C, \nu)$ denoting the Wishart distribution with scale matrix C and ν degrees of freedom, see Appendix D.3.

Proof. See Press [65], Theorems 7.1.2, 7.1.4 and 7.1.5. □

For general elliptical distributions we obtain the following distributional result for the maximum likelihood estimator for the mean.

Proposition 4.17. *Let $R_s \sim \mathcal{E}(\mu, \Sigma, \phi)$, $s = 1, \ldots, S$ i.i.d. Then, the maximum likelihood estimator $\hat{\mu}_S^{ML}$ has the following distribution:*

$$\hat{\mu}_S^{ML} \sim \mathcal{E}_n\left(\mu, \frac{1}{S}\Sigma, \phi^S\right)$$

with $\phi^S = \prod_{i=1}^{S} \phi$.

Proof. Follows from Proposition 4.10, part (v). □

Besides the maximum likelihood estimator there exist several other estimators for the mean of an elliptical distribution. An also quite well-known estimator is the median which represents an estimator based on empirical quantiles. As we are working with symmetric elliptical distributions, the marginal distributions are as well symmetric (see Proposition 4.10) and thus the median and the mean of the marginals coincide. Hence, using the empirical median as an estimator for the mean is meaningful. The same argument holds for the estimators presented thereafter, as they are all symmetrically built.

Before defining the median estimator, we need to introduce the notation for a (one-dimensional) ordered sample and the quantiles. Estimators based on quantiles will always be defined componentwise, i.e. on the respective marginals.

Definition 4.18. *The i-th component of each R_s, $s = 1, \ldots S$ represents a one-dimensional random sample, and by $R_{(1),i} \leq \ldots\ldots R_{(S),i}$ we denote the associated* ordered sample*. Then a point estimator for the theoretical α-quantile q_α ($0 < \alpha < 1$) of the according distribution is given by the* sample quantile

$$\mathcal{Q}_{q_\alpha,S,i} := R_{(\lfloor \alpha S \rfloor + 1),i}.$$

Having this definition of quantile estimators, the median estimator as the 50%-quantile is just a special case thereof:

Definition 4.19. *The* median estimator *or* sample median *for μ based on a sample of length S is defined componentwise by*

$$\hat{\mu}_{S,i}^{ME} := \mathcal{Q}_{q_{0.5},S,i} = R_{(\lfloor \frac{S}{2} \rfloor + 1),i}.$$

Analogously, further quantile-based estimators can be defined. In our case of symmetric distributions, an alternative to the median estimator is given generally by a (symmetrically) weighted mixture of some of its quantiles. We will choose in particular the average of the 25% and 75%-quartiles:

Definition 4.20. *The* quartile estimator *for μ based on a sample of length S is defined componentwise by*

$$\hat{\mu}_{S,i}^{QR} := \frac{1}{2}\mathcal{Q}_{q_{0.75},S,i} + \frac{1}{2}\mathcal{Q}_{q_{0.25},S,i} = \frac{1}{2}R_{(\lfloor 0.25\cdot S\rfloor+1),i} + \frac{1}{2}R_{(\lfloor 0.75\cdot S\rfloor+1),i}.$$

In the following we want to introduce two more estimators, originating in robust statistics: the trimmed mean and the Huber estimator. The α-trimmed mean estimator ignores the α percent smallest and largest values of the underlying sample and calculates the classical mean or average of the remaining sample.

Definition 4.21. *Let $0 < \alpha < \frac{1}{2}$. The* α-trimmed mean estimator *for μ based on a sample of length S is defined componentwise by*

$$\hat{\mu}_{S,i}^{TM} := \frac{1}{S - 2\lfloor \alpha S\rfloor} \sum_{s=\lfloor \alpha S\rfloor+1}^{S-\lfloor \alpha S\rfloor} R_{(s),i}.$$

All the estimators introduced so far are defined on the basis of an ordered sample. A general class of estimators is given by the so-called *L-estimates*, estimators that can be expressed as linear combinations of an ordered sample $R_{(1),i} \leq \ldots\ldots \leq R_{(S),i}$, i.e. as

$$\sum_{s=1}^{S} c_s R_{(s),i}$$

with appropriately chosen weights c_s, see Huber [38] for a proper general definition.

As all the previously presented estimators for the mean, i.e. the maximum likelihood estimator, the median, the quartile estimator and the trimmed mean estimator, can be represented by such a linear combination, they can all be subsumed within the class of L-estimates.

Remark 4.22. *For the four L-estimates introduced in this section, the weights c_s in the linear combination to define the general L-estimate have to be chosen as follows:*

- *Maximum likelihood estimator:*

$$c_s = \frac{1}{S} \quad \forall s = 1, \ldots, S.$$

- *Median:*

$$c_s = \begin{cases} 1 & \text{for } s = \lfloor \frac{S}{2}\rfloor + 1 \\ 0 & \text{otherwise.} \end{cases}$$

- *Quartile estimator:*

$$c_s = \begin{cases} \frac{1}{2} & \text{for } s = \lfloor 0.25 \cdot S \rfloor + 1 \text{ and } s = \lfloor 0.75 \cdot S \rfloor + 1 \\ 0 & \text{otherwise.} \end{cases}$$

- *α-trimmed mean estimator:*

$$c_s = \begin{cases} \frac{1}{S - 2\lfloor \alpha S \rfloor} & \text{for } s = \lfloor \alpha S \rfloor + 1, \ldots, S - \lfloor \alpha S \rfloor \\ 0 & \text{otherwise.} \end{cases}$$

The last estimator we want to present in this section is the Huber estimator, see e.g. Huber [37, 38]:

Definition 4.23. *The* Huber estimator $\hat{\mu}_S^{HU}$ *based on a sample of length S is defined componentwise by*

$$\hat{\mu}_{S,i}^{HU} := \arg\min_{y \in \mathbb{R}} \sum_{s=1}^{S} \rho(R_{s,i} - y) \quad \text{with} \quad \rho(x) = \begin{cases} \frac{x^2}{2} & \text{if } |x| \leq k \\ k|x| - \frac{k^2}{2} & \text{if } |x| > k \end{cases}$$

for some $k > 0$, see Huber [37].

Remark 4.24. *Depending on the choice of the parameter k, the Huber estimator does not necessarily have to be unique. Consider for example the case of a sample with the two (one-dimensional) points $R_1 = 1$ and $R_2 = -1$ and let $k = \frac{1}{2}$. Then the minimum is attained on the entire interval $[-\frac{1}{2}, \frac{1}{2}]$, i.e. the Huber estimator is not unique. For k sufficiently large, e.g.*

$$k \geq \frac{1}{2} \max_{1 \leq s \leq S-1} |R_{(s),i} - R_{(s+1),i}|,$$

i.e. k larger than half of the maximum distance of any two neighboring observations in the ordered sample, the Huber estimator is unique as it is attained on the interval where the function ρ is strictly convex. For more details, see the diploma thesis of Middelkamp [61], Remark 2.36.

Assumption 4.25. *In the following we assume that the Huber estimator is unique. In cases where k is not sufficiently large to assure uniqueness (e.g. if a constant k is fixed independent of the sample), we define the right end of the interval of optimal solutions to be* the *Huber estimator. Note that this particular choice is analogous to the definition of a unique median in Definition 4.19.*

Besides the class of L-estimates, there exists a further general classification of estimates, the so-called *M-estimates*. Those are maximum likelihood type

estimators minimizing some deviation measure ρ, i.e. for estimating a location parameter they can be expressed as

$$\arg\min_{z\in\mathbb{R}^n} \sum_{s=1}^{S} \rho(R_{s,i} - z).$$

From Definition 4.23 it is obvious that the Huber estimator falls within the class of M-estimates, but also the maximum likelihood estimator and the median can be expressed as a minimization problem with a suitable function ρ.

Remark 4.26. *The maximum likelihood and the median estimator both belong to the class of M-estimates:*

- *The MLE for μ can also be calculated by solving the following optimization problem for each component:*

$$\hat{\mu}_{S,i}^{ML} = \arg\min_{y\in\mathbb{R}^n} \sum_{s=1}^{S} (R_{s,i} - y)^2 = \arg\min_{y\in\mathbb{R}^n} \sum_{s=1}^{S} \rho(R_{s,i} - y)$$

  with $\rho(x) = x^2$.

- *The median estimator can as well be obtained by solving the following optimization problem for each component separately:*

$$\hat{\mu}_{S,i}^{ME} = \arg\min_{y\in\mathbb{R}} \sum_{s=1}^{S} |R_{s,i} - y| = \arg\min_{y\in\mathbb{R}} \sum_{s=1}^{S} \rho(R_{s,i} - y)$$

  with $\rho(x) = |x|$.
  Note that in case of an entire interval of optimal solutions, we again choose the right end to be the median.

As we want to use these presented estimates to approximate the mean vector μ in practical applications, we investigate if it theoretically matches the target parameter, i.e. we are interested in unbiased estimators. The following proposition summarizes the results.

Proposition 4.27. *Let $R_s \sim \mathcal{E}_n(\mu, \Sigma, \phi)$, $s = 1, \ldots, S$ i.i.d. Then it holds:*

(i) The estimators $\hat{\mu}_S^{ML}$, $\hat{\mu}_S^{ME}$, $\hat{\mu}_S^{QR}$, $\hat{\mu}_S^{TM}$ and $\hat{\mu}_S^{HU}$ as defined above are unbiased estimators for the vector μ.

(ii) In case of a normal distribution, the maximum likelihood estimator $\hat{\Sigma}_S^{ML}$ is a biased estimator for the covariance matrix Σ, but it is asymptotically unbiased.

Proof.

(i) Due to the symmetry of the marginal disributions of $R_s, s = 1, \ldots, S$, the first four estimators are unbiased as representatives of L-estimates according to Rinne [69], page 474. Furthermore, in our setting the Huber estimator is unbiased as well, as by Theorem 4 on page 364 in Goodall [34], M-estimates of location are unbiased.

(ii) In case of a normally distributed sample, the maximum likelihood estimator for the covariance follows a Wishart distribution (see e.g. Press [65], Theorem 7.1.5):

$$\hat{\Sigma}_S^{ML} \sim \mathcal{W}\left(\frac{1}{S}\Sigma, S-1\right).$$

The expectation of $\hat{\Sigma}_S^{ML}$ is then given by (see Proposition D.6)

$$\mathbf{E}[\hat{\Sigma}_S^{ML}] = \frac{S-1}{S}\Sigma,$$

hence, $\hat{\Sigma}_S^{ML}$ is biased, but asymptotically unbiased.

Note that the sample covariance matrix $\hat{\Sigma}_S^{SA} = \frac{S}{S-1}\hat{\Sigma}_S^{ML}$ is unbiased.

□

4.4 Portfolio theory and the classical optimization problem

In the preceding sections we have presented elliptical distributions that can be used to model a financial market, and we have discussed several parameter estimators that are used to determine the first two moments of an elliptically distributed random vector. In the following we introduce portfolios and their risk-return characteristics, define the efficient frontier and finally state the classical optimization problem which determines efficient portfolios. The foundation of portfolio optimization was laid by H. Markowitz [56] in the 1950s when he introduced the concept of mean-variance analysis. The basic idea is that the mean and the variance are the only quantities to characterize the risk-return-profile of a portfolio.

Consider a financial market of n risky assets and assume that the vector of their returns, $R \in \mathbb{R}^n$ is distributed according to an elliptical distribution with mean μ and covariance matrix Σ, i.e. the vector of expected asset returns is given by

$$\mathbf{E}[R] = \mu$$

and the covariance matrix describing the relation between the individual asset returns is

$$\mathbf{Cov}[R] = \Sigma.$$

Very often in practical applications, it is assumed that the returns follow a multivariate normal distribution, i.e.

$$R \sim \mathcal{N}(\mu, \Sigma).$$

Based upon such a market, we want to determine *optimal portfolios* where "optimal" is characterized only by the *risk* and the *return* of a portfolio, measured by its expectation and variance. Thus, naturally, the resulting optimization problem relies on the market parameters μ and Σ. Generally, we would like to use the particular parameter values that will really represent the market in the (future) period of our investment, e.g. if investing today for the time period of a month, we would like to know the expected performance of each asset from now until then to be in the advantageous position to make the optimal investment decision today. But as those values are naturally unknown, estimates have to be used instead.

Let the vector $x = (x_1, \ldots, x_n)^T$ with $x^T\mathbf{1} = 1$ denote the weights of a portfolio, i.e. the proportional investments in the n risky assets. Let furthermore $R \in \mathbb{R}^n$ describe the vector of asset returns in the considered period of time[4]. The return of the portfolio x is then given by

$$R(x) = x^T R.$$

The expected portfolio return is accordingly

$$\mu(x) := \mathbf{E}[R(x)] = x^T\,\mathbf{E}[R] = x^T\mu.$$

In the mean-variance framework the risk of a portfolio is measured by the variance (or equivalently by the standard deviation), i.e. the risk of portfolio x is given by

$$\sigma^2(x) := \mathbf{Cov}[x^T R] = x^T\,\mathbf{Cov}[R]x = x^T\Sigma x.$$

There exists quite a bit of literature mistrusting the use of the variance as an appropriate risk measure and postulating other ones. Criticism mostly refers to the equal handling of upside and downside deviations. Whereas upside deviations (i.e. the resulting portfolio return is larger than expected) are welcome and do not represent a risk in practice, downside deviations can imply rather severe losses

[4]The Markowitz setting represents a one-period investigation of a portfolio, i.e. no dynamics or changing allocations over time are considered in this framework. Rather common time periods are e.g. weeks, months or years, and the calibration of the parameters for the optimization is usually based on historical data with the same frequency.

and consequences. A further drawback of the variance as a risk measure is that many practitioners have other needs, meaning that controlling the variance (or the standard deviation) is not sufficient to express the investor's risk. Very common in financial practice is to use the Value-at-Risk (VaR) or the Conditional Value-at-Risk (CVaR) as comparison figure, or other closely related shortfall measures. The VaR expresses the maximum loss an investor can encounter in a given period of time and with prescribed probability. Shortfall measures are for example used in applications where it is rather unimportant how much the final value deviates from the expectation as long as it does not fall below a certain critical value. Such a risk could be measured by the shortfall probability, the probability of being below a given benchmark at the end of the period. Additionally, the average loss in case of falling below the benchmark is an interesting quantity in practical applications that might be wanted to be controlled. In case of describing the benchmark by the VaR, this measure is called the CVaR.

For further discussions of various risk measures in financial practice and applications thereof, we refer to the literature. Properties of general risk measures and their role in optimization are investigated by Rockafellar et al. in [73] and [71], Rockafellar and Uryasev also study especially the VaR and the CVaR in [72]. Of interest when dealing with risk measures is the characteristic of a coherent risk measure, a desired property introduced by Artzner et al. [2].

Despite all drawbacks and discussions, using the variance as a risk measure is still very popular. And since in many applications a multivariate normal distribution (or more generally an elliptically symmetric distribution) is assumed for the return vectors, mean-variance analysis is appropriate, see [63].

Before finally introducing the portfolio optimization problem, we need to define the notion of an *efficient portfolio* and the *efficient frontier*.

Definition 4.28. *Let $x, y \in \mathbb{R}^n$ with $x \neq y$ denote two different portfolio allocations.*

- *Portfolio x is said to* dominate *portfolio y if it has a higher (or equal) expected return and a smaller (or equal) risk, i.e. if*

$$\mu(x) \geq \mu(y) \quad \text{and} \quad \sigma^2(x) \leq \sigma^2(y).$$

- *Portfolio x is said to* strictly dominate *portfolio y, if at least one of the inequalities is strict.*

- *A portfolio x is called* efficient *if there does not exist a portfolio y that strictly dominates x.*

Equivalently, it can be said that a portfolio is efficient if it yields the highest expected return to a given risk level, or if it has the smallest variance to a fixed level of expected return.

Definition 4.29. *The* efficient frontier *is described by the set*

$$\left\{ \big(\mu(x), \sigma(x)\big) \mid x \text{ efficient} \right\},$$

i.e. the efficient frontier is is the curve in a risk-return diagram representing the characteristics of all efficient portfolios.

Naturally, the aim is to find efficient portfolios. But, as can easily be observed by the definition of an efficient portfolio, there are two competing objectives: minimizing the risk of the portfolio (i.e. the variance or the standard deviation) and maximizing the expected return. Thus, a trade-off between risk and return has to be made.

Throughout the dissertation we will consider the portfolio optimization problem for determining optimal, i.e. efficient, portfolios in the following form:

Definition 4.30. *The classical portfolio optimization problem is given by*

$$\min_{x \in X} \quad (1-\lambda)\sqrt{x^T \Sigma x} - \lambda x^T \mu \tag{P_λ}$$

with $X \subset \{x \in \mathbb{R}^n \mid x^T \mathbf{1} = 1\}$ *non-empty, convex and compact, and* $\mu \in \mathbb{R}^n$ *and* $\Sigma \in \mathbb{R}^{n \times n}, \Sigma \succ 0$ *describing the expected return and the covariance matrix of the asset returns. The parameter* $\lambda \in [0,1]$ *expresses the relation (or the trade-off) between risk and return of the portfolio. The optimal solution*[5] *to* (P_λ) *for a given trade-off parameter* λ *is denoted by* $x^*_{cl}(\lambda)$*.*

Thus, for each value of trade-off parameter λ an efficient portfolio is determined by the program (P_λ), and by letting λ increase from zero to one, we therefore trace the entire efficient frontier. The outmost portfolios represent two prominent ones:

(i) For $\lambda = 0$ the optimization problem (P_λ) reduces to

$$\min_{x \in X} \quad \sqrt{x^T \Sigma x}$$

which finds the portfolio with the lowest risk while not incorporating any information about expected returns. This particular portfolio is called the *minimum variance portfolio* (MVP) and defines the left end of the efficient frontier.

(ii) For $\lambda = 1$ program (P_λ) simplifies to

$$\max_{x \in X} \quad x^T \mu.$$

In this case the objective is to maximize the expected return that can be achieved by any feasible portfolio, omitting any risk considerations. This portfolio denotes the right end of the efficient frontier and is called *maximum return portfolio* (MRP). Note that this problem was already investigated from a theoretical point of view in Section 2.3.6 and Example 3.27.

[5]For uniqueness of the optimal solution see Proposition 4.31 below.

The feasibility set X is supposed to be non-empty, convex and compact containing at least the condition that makes $x \in X$ a portfolio, meaning that the sum of of the components of x has to be 1, i.e. $X \subset \{x \in \mathbb{R}^n \mid x^T\mathbf{1} = 1\}$. The set X possibly contains further constraints on the asset weights that do not depend on the (uncertain) parameters μ and Σ. Common definitions of X in financial applications are the following:

- $X = \{x \in \mathbb{R}^n \mid x^T\mathbf{1} = 1\}$ with $\mathbf{1}$ denoting the vector consisting of 1 in each component. This is the minimum set to make the variable x represent a portfolio. Note that this feasibility set is not compact.
- In most applications shortselling is not allowed, i.e. the portfolio x must not have any negative entries. Thus, a rather popular set of constraints on x is given by the compact set $X = \{x \in \mathbb{R}^n \mid x^T\mathbf{1} = 1, x \geq 0\}$.
- In practice there exist very often constraints limiting the investment in a particular asset or in all assets (e.g. in no individual asset may be invested more than 10%), or in a set of assets (e.g. the investment in equities may at most be 30%). Such linear constraints can be summarized in a feasibility set of the form $X = \{x \in \mathbb{R}^n \mid x^T\mathbf{1} = 1, Ax \leq b\}$ with A and b such that X is non-empty, compact and convex. Note that x being non-negative can also be incorporated into the linear inequalities.

The following proposition shows that the portfolio optimization problem (P_λ) has a unique optimal solution for $0 \leq \lambda < 1$. According to Proposition 2.45 uniqueness thus implies continuity with respect to the parameters, a result that will be needed again in Section 6.3 to prove consistency of the optimal portfolios.

Proposition 4.31. *Let $0 \leq \lambda < 1$. Then program (P_λ) as given in Definition 4.30 has a unique optimal solution $x^*(\lambda)$.*

Proof. We first consider the case $\lambda = 0$. Problem (P_0) thus reduces to

$$\min_{x \in X} \sqrt{x^T \Sigma x}.$$

As $\sqrt{x^T\Sigma x} \geq 0$ and the function $h(z) = z^2$ is a strictly increasing function for positive z, this program is equivalent to the problem with squared objective function:

$$\min_{x \in X} \quad x^T \Sigma x$$

which has a unique optimal solution $x^*(0)$ since the objective function is strictly convex (recall that the covariance matrix Σ is positive definite).

Let $0 < \lambda < 1$. We want to show uniqueness of the optimal solution by proving that the objective function is strictly convex over the set of feasible x,

i.e. over X. Note that it always holds that $X \subset \{x \in \mathbb{R}^n \mid x^T\mathbf{1} = 1\}$. To show strict convexity of the objective function on X, it suffices to analyze the term

$$h(x) := \sqrt{x^T \Sigma x} = \|Ax\|$$

with $A = \Sigma^{\frac{1}{2}} \succ 0$. Since the feasibility set X is convex, it holds (see e.g. Boyd and Vandenberghe [19], Section 3.1) that the function h is strictly convex on X if and only if

$$h(y) > h(x) + \nabla h(x)^T (y - x) \quad \forall x, y \in X. \tag{4.4}$$

Since $h(x) = \|Ax\|$ and $\|Ax\| > 0$ for all $x \in X$, the gradient is given by

$$\nabla h(x) = \frac{A^T A x}{\|Ax\|}$$

as already calculated in Theorem 3.28. To prove Equation 4.4, we start with the Cauchy-Schwarz inequality:

$$|x^T y| \leq \|x\| \, \|y\|$$

which holds with equality if and only if $y = k \cdot x$, see e.g. [74], Theorem 9.2. Hence, in our setting with $x, y \in X \subset \{x \in \mathbb{R}^n \mid x^T\mathbf{1} = 1\}$ and $x \neq y$ the Cauchy-Schwarz inequality is strict, i.e. it holds

$$|x^T y| < \|x\| \, \|y\| \quad \forall x, y \in X, \; x \neq y.$$

As $A = \Sigma^{\frac{1}{2}}$ is invertible, $y = k \cdot x$ is equivalent to $Ay = k \cdot Ax$ and we also have the strict Cauchy-Schwarz inequality for the vectors Ax and Ay:

$$|(Ax)^T (Ay)| < \|Ax\| \, \|Ay\| \quad \forall x, y \in X, \; x \neq y.$$

Using that $(Ax)^T (Ay) \leq |(Ax)^T (Ay)|$ and $\|Ax\| > 0$ (as $x \neq 0$), it follows that

$$\begin{aligned} & \frac{x^T A^T A y}{\|Ax\|} < \|Ay\| \\ \Leftrightarrow \quad & \frac{x^T A^T A (y - x)}{\|Ax\|} < \|Ay\| - \underbrace{\frac{x^T A^T A x}{\|Ax\|}}_{= \|Ax\|} \end{aligned}$$

and hence finally gives

$$\|Ay\| > \|Ax\| + \left(\frac{A^T A x}{\|Ax\|} \right)^T (y - x) \quad \forall x, y \in X, \; x \neq y.$$

Hence, for $h(x) = \|Ax\|$ and $x, y \in X$ arbitrary, the strict inequality (4.4) holds and thus $h(x)$ is strictly convex over X. Finally, minimizing a strictly convex function over a compact set yields a unique solution $x^*(\lambda)$. □

Remark 4.32. *Alternatively, Theorem 3.28 could be used to show uniqueness of the optimal solution of (P_λ) for $0 < \lambda < 1$, as the objective function can be interpreted as a robustified optimization problem: Consider the program*

$$\min_{x \in X} \quad -x^T r \tag{P_{aux}}$$

with an ellipsoidal uncertainty set around the parameter μ described by

$$\mathcal{U}_\delta(\mu) = \{r \in \mathbb{R}^n \mid (r-\mu)^T \Sigma^{-1} (r-\mu) \leq \delta^2\}$$

with $\delta := \frac{1-\lambda}{\lambda}, 0 < \lambda < 1$. Hence, the according robust optimization problem results in the following (for explicit reformulations see Example 3.26):

$$\begin{aligned} &\min_{x \in X} \quad -x^T\mu + \delta \|\Sigma^{\frac{1}{2}} x\| \\ =&\min_{x \in X} \quad \frac{1-\lambda}{\lambda} \sqrt{x^T \Sigma x} - x^T \mu \end{aligned}$$

which, for $\lambda > 0$, is equivalent to

$$\min_{x \in X} \quad (1-\lambda)\sqrt{x^T \Sigma x} - \lambda x^T \mu,$$

the problem under consideration. As (P_{aux}) fulfills the prerequisites of Theorem 3.28 and especially the consequences thereafter, we can conclude that the optimal solution of (P_λ) with $0 < \lambda < 1$ is unique.

Note that uniqueness of the optimal solution also guarantees continuity of the optimal solution with respect to the uncertain parameters μ and Σ (see Theorem 2.45), since the feasibility set X is constant and thus Hausdorff continuous.

Remark 4.33. *In case of $\lambda = 1$, problem (P_1) simplifies to*

$$\min_{x \in X} \quad -x^T \mu$$

which does not guarantee a unique solution in its general form. In case of the feasibility set X being described by

$$X = \{x \in \mathbb{R}^n \mid x^T \mathbf{1} = 1, x \geq 0\},$$

a simple but rather common constraint set, the solution of the optimization problem is given by the portfolio investing 100% in the asset with the highest expected return and nothing in any other asset. Hence, it then suffices to assume that the maximum component of the vector μ is unique to assure uniqueness of the optimal solution as well in case of $\lambda = 1$.

It is rather well known (see e.g. Jorion [48] or Best and Grauer [13]) that the two parameters μ and Σ have a very large influence on the optimal solution; especially the vector of expected return determines the optimal allocation quite heavily, and rather small changes in the returns can imply drastic changes in the optimal portfolio. A very illustrative example for this fact is the following simple example of the maximum return portfolio which we have studied several times already.

Example 4.34. *Consider a market that only consists of two risky assets, and we want to determine the maximum return portfolio, i.e. we want to solve*

$$\max_{x \in X} \quad x^T \mu.$$

The set of constraints is supposed to be described by

$$X = \{x \in \mathbb{R}^n \mid x^T \mathbf{1} = 1, x \geq 0\}.$$

Note that this is the same problem as in Section 2.3.6 and Example 3.27.

Now let us assume that a first parameter estimate for μ is given by the vector $\hat{\mu}_A = (5\%, 5.1\%)^T$. As the maximum return portfolio does not account for any risk, it simply chooses the asset with the highest expected return, thus, the optimal solution in this case is given by $x_A^* = (0, 1)^T$. On the other hand, having a parameter estimate of $\hat{\mu}_B = (5\%, 4.9\%)^T$ – such a small change in estimators can easily happen, for example by just having a few more historical observations – the optimal solution turns out to be $x_B^* = (1, 0)^T$. Thus, a small change in the parameter estimate (small enough to be considered simply as an estimation error) can completely turn the portfolio allocation.

Note that the objective value in both cases is quite similar. As we know from Proposition 2.40 or Section 2.3.6, the optimal objective value is continuous in the (uncertain) data, hence the portfolio return is only marginally affected by minor changes in the parameter. But as in this particular example the optimal solution (i.e. the portfolio allocation) is not continuous in the data, such an extreme change can occur.

In all the subsequent examples and plots, we will restrict the feasible portfolios to consisting of long-only positions, i.e. we assume the feasibility set X to be given by

$$X = \{x \in \mathbb{R}^n \mid x^T \mathbf{1} = 1, x \geq 0\}.$$

In the following we illustrate the influence of the parameters on the optimization result in the extreme case of the maximum return portfolio over time, i.e. moving along the time axis, calculating in every point the corresponding maximum likelihood estimators and solving the optimization problem for $\lambda = 1$. This results in the following plot (Figure 4.6), showing the allocation of the maximum return portfolio at each time point. As we have imposed the long-only constraint, the

MRP always consists of only one asset – the one with the highest[6] expected return. It can nicely be seen that in the period where the bear market data are

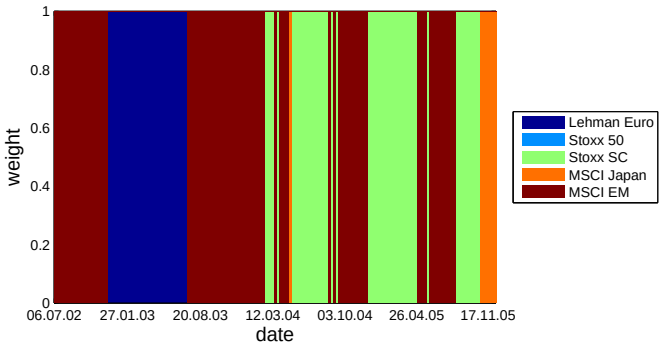


Figure 4.6: Allocation of the MRP over time.

used to estimate the parameters (roughly the first half of the year 2003), the bond index Lehman Euro has the highest expected return and thus forms the maximum return portfolio. In other market phases one of the stock indices has the best expected performance. Roughly in September 2004 there are several alternating changes between two assets having the highest expected return. Such a behavior is rather common and suggests that these assets approximately have the same return and that robustification could prevent the alternations.

For the minimum variance portfolio (MVP) at the other end of the efficient frontier, the situation is different. The MVP does not depend at all on the vector of expected returns μ, but only on the covariance matrix Σ which influences the result of the optimization problem not as significantly as the vector of expected returns. Hence, the allocation of the minimum variance portfolio does not exhibit such an extreme behavior as the maximum return portfolio, but is rather stable, see Figure 4.7, as it is always invested to more than 80% in the bond index – which is much less volatile than the stock indices. The remaining part of the asset allocation of the MVP nevertheless exhibits a little variation over time.

To illustrate the changes for a portfolio somewhere in between, we choose the maximum sharpe ratio portfolio (MSRP), i.e. the portfolio maximizing the sharpe ratio SR, defined as

$$SR = \frac{\mu(x) - r_0}{\sigma(x)}$$

[6]The probability of two or more assets having exactly the same highest return is almost surely zero when estimating the parameters from a finite sample of realizations.

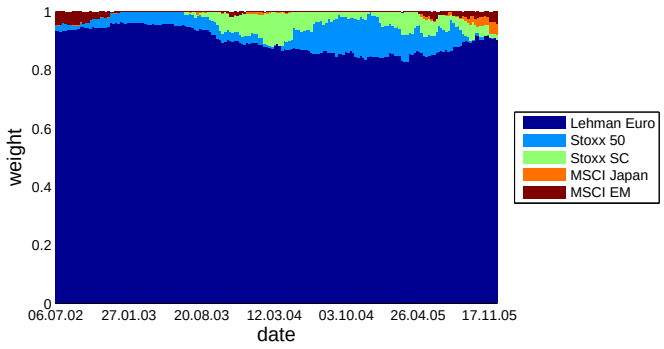


Figure 4.7: Allocation of the MVP over time.

with $\mu(x)$ and $\sigma(x)$ denoting the expected return and volatility of the portfolio x, and r_0 being the riskless interest rate which was set to 2% p.a. in the calculations for simplicity. It is worth noting that the maximum sharpe ratio portfolio does not correspond to a particular value of λ, but it can be placed very differently on the efficient frontier. This is shown in Figure 4.8 which plots the value of lambda yielding the maximum sharpe ratio portfolio at each point in time. From

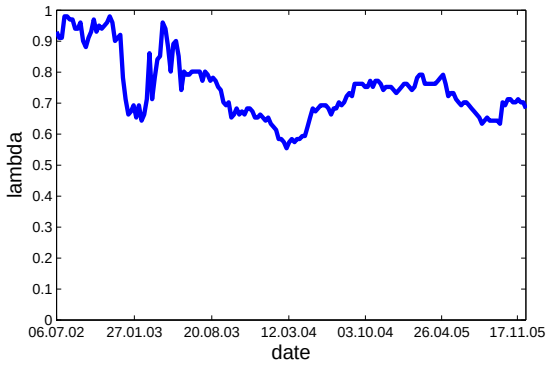


Figure 4.8: The values of λ corresponding to the MSRPs.

Figure 4.9 it can be observed that the MSRP has more changes in the allocation than the minimum variance portfolio, but is not as extreme as the maximum return portfolio. At the beginning of the time period, the maximum sharpe ratio portfolio resembles the maximum return portfolio. Sometimes it is invested in both the emerging market index and the bond index, but often it is even equal

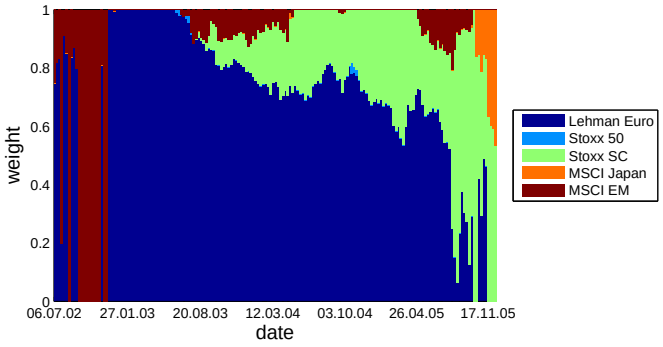


Figure 4.9: Allocation of the MSRP over time.

to the MRP, especially in the market phase where the bear market determined the parameters and the best perfoming asset was the Lehman Euro.

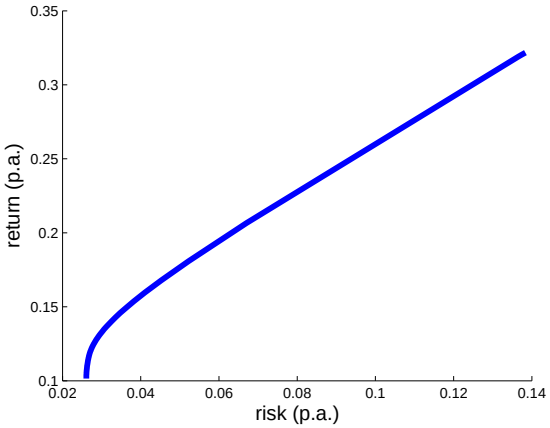
So far we have fixed a particular portfolio and monitored its changes along the time axis. Next we illustrate the changes in the portfolio allocation along the efficient frontier and not over time. Hence, we pick a point in time (here we arbitrarily choose the 01.11.2003) and calculate the corresponding maximum likelihood estimators which are summarized in Table 4.3. Instead of the covariance matrix which is needed in the optimization problem, the individual volatilities of the assets and their correlation matrix are stated for easier interpretation.

	return	volatility	correlation matrix				
Lehman Eur	9.2%	3.1%	1.00	-0.41	-0.36	-0.09	-0.21
Stoxx 50	5.9%	22.1%	-0.41	1.00	0.80	0.29	0.60
Stoxx SC	27.0%	14.6%	-0.36	0.80	1.00	0.50	0.70
MSCI Japan	19.0%	19.5%	-0.09	0.29	0.50	1.00	0.57
MSCI EM	32.2%	13.9%	-0.21	0.60	0.70	0.57	1.00

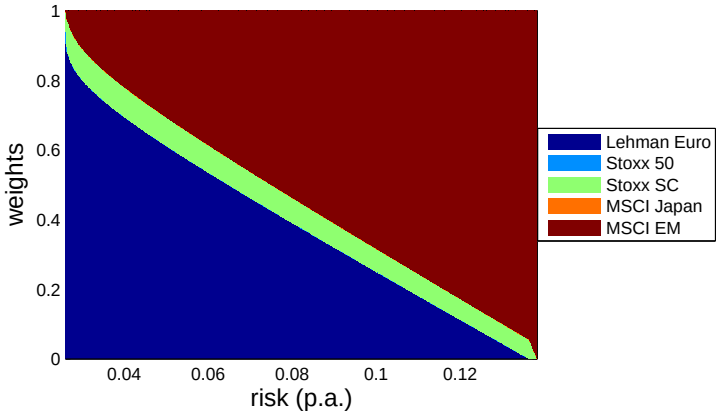
Table 4.3: Annualized (multiplied by 52 resp. $\sqrt{52}$) asset returns and volatilities and the correlation on 01.11.2003.

Using the estimated parameters, the efficient portfolios are obtained by solving the problem (P_λ) for $0 \leq \lambda \leq 1$. The respective efficient frontiers (again plottet with annualized values) and the associated changing portfolio allocations along the efficient frontier are shown in Figure 4.10.

From the allocation picture it can be observed that the minimum variance portfolio, determining the left end of the efficient frontier, consists to a very large amount of the Lehman Euro, the only bond index in our asset universe. As the



(a) efficient frontier



(b) portfolio allocations

Figure 4.10: Illustration of the classical efficient frontier and the associated portfolio allocations for the time point 01.11.2003.

bond has a substantially smaller volatility than all the stocks at this point of time, this is not surprising. The more risky the portfolios get, i.e. moving along the efficient frontier to the right, the smaller is the portion that is invested in the bond index, as this asset does not yield as high returns as the more volatile stock indices. The Lehman Euro and the Emerging Market Index had the best Sharpe ratio (larger than 2) among the five assets, and hence the largest part of the portfolio was invested in those two indices, smoothly adapted with increasing risk tolerance. It can further be seen again in the figure that the maximum return portfolio consists only of one asset – the Emerging Market Index in this case.

As the strong dependence of the Markowitz solution on the input data is known (see e.g. [13], [21] or [48]), solutions for this drawback are sought for. Besides using more robust parameter estimates in the optimization (see e.g. [25],

[45] or [47]) or resampling the procedure for obtaining efficient frontiers (see e.g. [48] or [60]), the robust counterpart approach as presented in its general form in Chapter 3 can be applied to the traditional portfolio optimization problem (P_λ) to explicitly account for uncertainty in the estimation.

Chapter 5

Robust portfolio optimization

5.1 The robust portfolio optimization problem

In Section 4.4 we have introduced the portfolio optimization problem in its traditional form and we have illustrated the strong dependence of its optimal solution on the input parameters μ and Σ. We now define the associated robust counterpart formulation of the classical problem which we recall for completeness:

$$\min_{x \in X} \quad (1-\lambda)\sqrt{x^T \Sigma x} - \lambda x^T \mu \tag{P_λ}$$

with $X \subset \{x \in \mathbb{R}^n \mid x^T \mathbf{1} = 1\}$ non-empty, convex and compact, and $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}, \Sigma \succ 0$ representing the expected return and the covariance matrix of the asset returns.

As both the vector of expected asset returns μ and the covariance matrix Σ are considered to be exposed to variability, the uncertain parameter u from the general convex optimization program (GCP_u) represents the pair (μ, Σ). In practical problems, there is often only defined an uncertainty set for the vector of expected returns, as the covariance matrix is not as volatile and furthermore does not as crucially affect the optimal solution as the return estimate (see e.g. [21]). Recall also Figure 4.3 for an illustration of this fact. In the general description of the robust optimization problem, we will nevertheless consider the covariance matrix as an uncertain parameter. Applying the robust counterpart approach to the portfolio optimization problem (P_λ) results in the following robustified program.

Definition 5.1. *The general form of the robust counterpart to problem (P_λ) is given by*

$$\min_{x \in X} \max_{(r,C) \in \mathcal{U}} \quad (1-\lambda)\sqrt{x^T C x} - \lambda x^T r \tag{RP_λ}$$

with $\mathcal{U}$ being the (joint) uncertainty set for the unknown parameters (μ, Σ). Analogous to the classical setting, the optimal solution of this robust problem will be denoted by $x^*_{rob}(\lambda)$.

The following proposition shows that – analogous to the classical portfolio optimization problem – uniqueness of the robust optimal solution can be assured for $0 \leq \lambda < 1$. The proof is based on the result of Lemma 3.17 that strict convexity of the classical objective function transfers to the robust objective. As in the classical case, uniqueness of the optimal solution implies continuity thereof with respect to the parameters.

Proposition 5.2. *Let the robust portfolio optimization problem (RP_λ) be given with $X \subset \{x \in \mathbb{R}^n \mid x^T\mathbf{1} = 1\}$. Assume furthermore an ellipsoidal uncertainty set for (μ, Σ), i.e. consider the problem*

$$\min_{x \in X} \max_{(r,C) \in \mathcal{U}(\mu,\Sigma)} \quad (1-\lambda)\sqrt{x^T C x} - \lambda x^T r.$$

Then the optimal solution $x^*_{rob}(\lambda)$ is unique for $0 \leq \lambda < 1$.

Proof. In the proof of Proposition 4.31 we have shown that for $0 \leq \lambda < 1$ the objective function of the classical portfolio optimization problem is strictly convex over the feasibility set X. Lemma 3.17, part (iii) hence gives strict convexity of the robust objective function, and hence $x^*_{rob}(\lambda)$ is unique for $0 \leq \lambda < 1$. □

Remark 5.3. *In case of $\lambda = 1$, the optimal solution $x^*_{rob}(1)$ is unique according to Corollary 3.32 if an ellipsoidal uncertainty set for μ (note that Σ is not needed for determining the maximum return portfolio) with full rank is employed.*

Notation 5.4. *In case of an uncertainty set only for the return, the formal definition of a joint uncertainty set for both μ and Σ centered at the point estimates $(\hat{\mu}, \hat{\Sigma})$ reduces to*

$$\mathcal{U}(\hat{\mu}, \hat{\Sigma}) = \mathcal{U}(\hat{\mu}) \times \{\hat{\Sigma}\}.$$

For ease of notation we will often neglect the part belonging to $\hat{\Sigma}$ and simply use the shorter expression $\mathcal{U}(\hat{\mu})$ while plugging in $\hat{\Sigma}$ directly into the formula. Note that compared to Chapter 3 we also omitted the subscript δ, as we are not particularly interested in properties referring to the size of the respective uncertainty sets. We will only include the size explicitly where appropriate.

For reformulating the robust optimization problem, an explicit uncertainty set has to be specified which represents the practical needs and is simple enough so that the resulting optimization problem can eventually be solved.

In the literature the robust counterpart is applied to the portfolio optimization problem with different uncertainty sets. Goldfarb and Iyengar [33] use the robust counterpart to model robust asset returns with a factor model and assume interval uncertainty for the mean and ellipsoidal uncertainty for the matrix of the factor loadings. Tütüncü and Koenig [79] prefer interval uncertainty sets where the endpoints could be determined from extreme values of e.g. historical data.

In the approach to finding the worst-case Value-at-Risk of El-Ghaoui, Oks and Oustry [27], they consider both a polytope uncertainty set and the case where the components of the mean and the covariance are supposed to lie componentwise within given bounds. Ben-Tal, Nemirovski and Margalit [6] illustrate multi-period portfolio optimization using an ellipsoidal uncertainty set and Lobo [52] investigates box and ellipsoidal uncertainty sets for the mean and the entries of the covariance matrix. Ellipsoidal uncertainty based on a confidence ellipsoid is as well used by Lutgens [55] to solve differently formulated portfolio optimization problems in the mean-variance framework.

We will in the following sections study explicit uncertainty sets for solving the robust formulation of the Markowitz portfolio optimization problem. As we have already seen (see Theorem 3.28) that an ellipsoidal uncertainty set is more promising than a polyhedral one, we will in the following always create uncertainty sets with ellipsoidal shape. Section 5.2 analyzes the idea of using a confidence ellipsoid around a point estimate to define an appropriate uncertainty set. We illustrate that such an approach can be used to create an uncertainty set only for μ or to define a joint uncertainty set for (μ, Σ). Section 5.3 makes use of different statistical estimators as presented in Section 4.3 to determine a practical uncertainty set for the vector of expected returns.

The optimal solution of the classical portfolio optimization problem (P_λ) for a given trade-off value λ will be denoted by $x^*_{cl}(\lambda)$ and the robust optimal solution is analogously given by $x^*_{rob}(\lambda)$. Note that here we have implicitly made the assumption of the optimal solutions being unique for each $\lambda \in [0, 1]$. As this was proved in the classical setting for $\lambda \in [0, 1)$ in Proposition 4.31, the assumption reduces to the following:

Assumption 5.5. *Let the classical maximum return portfolio be unique, i.e. the set of optimal solutions to problem (P_1) is a singleton.*

As we will only consider non-degenerate matrices to define ellipsoidal uncertainty sets in this chapter, the robust optimal solution $x^*_{rob}(\lambda)$ is unique for $\lambda \in [0, 1]$ according to Proposition 5.2 and Remark 5.3.

5.2 Confidence ellipsoid around the MLE

Defining an uncertainty set via the classical confidence ellipsoid around a point estimate is a rather intuitive and natural method coming from the field of stochastics. The center is given by the respective point estimate, the shape is described by an according covariance matrix (since for an elliptical distribution the covariance matrix determines the shape of the level curves around the peak, i.e. the mean), and the size of the uncertainty set is determined by the desired level of confidence.

In the following subsections we distiniguish the cases of creating a confidence ellipsoid around the maximum likelihood estimator for the mean vector μ only and a confidence ellipsoid jointly for μ and Σ.

5.2.1 Confidence ellipsoid for μ

To form an uncertainty set around the mean vector μ – or, more precisely, around an estimate $\hat{\mu}$ for the mean vector since the true market parameter is unknown – we need the distribution thereof. The distribution of the maximum likelihood estimator $\hat{\mu}^{ML}$ based on an i.i.d. sample $R_s \sim \mathcal{E}(\hat{\mu}, \hat{\Sigma}, \phi)$, $s = 1, \ldots, S$ is given by, see Proposition 4.17,

$$\hat{\mu}^{ML} \sim \mathcal{E}_n\left(\hat{\mu}, \frac{1}{S}\hat{\Sigma}, \phi^S\right)$$

with $\phi^S = \prod_{i=1}^{S} \phi$. As for elliptical distributions knowledge of the first two moments suffices for the definition of the confidence ellipsoid, we can thus create a confidence ellipsoid for the MLE centered at the point estimate $\hat{\mu}$ and using $\hat{\Sigma}$ to describe the shape:

$$\begin{aligned}
\mathcal{U}(\hat{\mu}) &= \{\mu \in \mathbb{R}^n \mid (\mu - \mathbf{E}[\hat{\mu}^{ML}])^T \left(\mathbf{Cov}[\hat{\mu}^{ML}]\right)^{-1} (\mu - \mathbf{E}[\hat{\mu}^{ML}]) \leq \delta^2\} \\
&= \{\mu \in \mathbb{R}^n \mid (\mu - \hat{\mu})^T \left(\frac{1}{S}\hat{\Sigma}\right)^{-1} (\mu - \hat{\mu}) \leq \delta^2\} \\
&= \left\{\mu \in \mathbb{R}^n \mid (\mu - \hat{\mu})^T \hat{\Sigma}^{-1} (\mu - \hat{\mu}) \leq \frac{\delta^2}{S}\right\} \qquad (5.1)
\end{aligned}$$

where the size δ^2 is determined by the desired confidence. It is known (see e.g. Anderson [1], Theorem 3.3.3) that in case of $R \sim \mathcal{N}(\hat{\mu}, \hat{\Sigma})$ the expression $(R - \hat{\mu})^T \hat{\Sigma}^{-1} (R - \hat{\mu})$ follows a χ^2 distribution with n degrees of freedom. Thus, the size δ^2 can be obtained by an appropriate α-quantile, i.e. δ^2 such that $\alpha = \chi_n^2(\delta^2)$, with $\alpha \in (0, 1)$ representing the confidence. Figure 5.1 illustrates in a two-dimensional example uncertainty sets originating in confidence ellipoids for different values of the confidence level α.

Remark 5.6. *Note that from this figure we could also deduce that the Lehman Euro bond index has a smaller volatility than the Stoxx 50 as the respective axis of the ellipse is shorter. Furthermore, as the ellipse is almost parallel to the coordinate axes, the correlation between the two assets is not too high.*

Using the uncertainty set from Equation (5.1), the worst case parameter μ_{wc} is obtained by solving

$$\begin{aligned}
&\max_{\mu \in \mathcal{U}(\hat{\mu})} \quad (1 - \lambda)\sqrt{x^T \hat{\Sigma} x} - \lambda(x^T \mu) && (5.2) \\
\Leftrightarrow \quad &\min_{\mu \in \mathcal{U}(\hat{\mu})} \quad \lambda x^T \mu && (5.3)
\end{aligned}$$

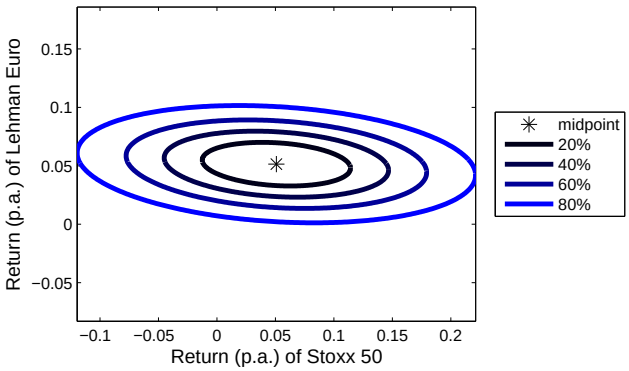


Figure 5.1: Confidence ellipsoids for two assets.

and is thus given by (see Example 3.26)

$$\mu_{wc} = \hat{\mu} - \frac{\delta}{\sqrt{S}} \frac{1}{\sqrt{x^T \hat{\Sigma} x}} \hat{\Sigma} x. \tag{5.4}$$

Incorporating this μ_{wc}, the robust counterpart problem becomes

$$\begin{aligned} &\min_{x \in X} \max_{\mu \in \mathcal{U}(\hat{\mu})} \quad (1-\lambda)\sqrt{x^T \hat{\Sigma} x} - \lambda(x^T \mu) \\ =&\min_{x \in X} \quad (1-\lambda)\sqrt{x^T \hat{\Sigma} x} - \lambda(x^T \hat{\mu}) + \lambda \frac{\delta}{\sqrt{S}} \sqrt{x^T \hat{\Sigma} x} \\ =&\min_{x \in X} \quad \left(1 - \lambda + \lambda \frac{\delta}{\sqrt{S}}\right) \sqrt{x^T \hat{\Sigma} x} - \lambda(x^T \hat{\mu}). \qquad (RP_{\lambda,\text{conf}}) \end{aligned}$$

In this particular – but nonetheless well-known and often used – setting, a rather surprising result can be found: the robust efficient frontier equals the classical efficient frontier, except that it is "shortened" with respect to the risk axis, i.e. it does not reach portfolios with as high risk as the maximum return portfolio in the classical framework. This result is illustrated in Figure 5.2 and stated formally in the following Proposition 5.7.

Proposition 5.7. *Consider the portfolio optimization problem (P_λ) and let the uncertainty set for the parameter vector μ be given by a confidence ellipsoid around the MLE as described in Formula (5.1). Then for each $\theta \in [0,1]$ there exists a $\lambda \in \left[0, \frac{1}{1+\frac{\delta}{\sqrt{S}}}\right]$, $\lambda = \lambda(\theta) = \frac{\theta}{1+\theta\frac{\delta}{\sqrt{S}}}$, such that the optimal solution $x^*_{rob}(\theta)$ of the corresponding local robust counterpart problem equals the optimal solution $x^*_{cl}(\lambda)$ of the original problem.*

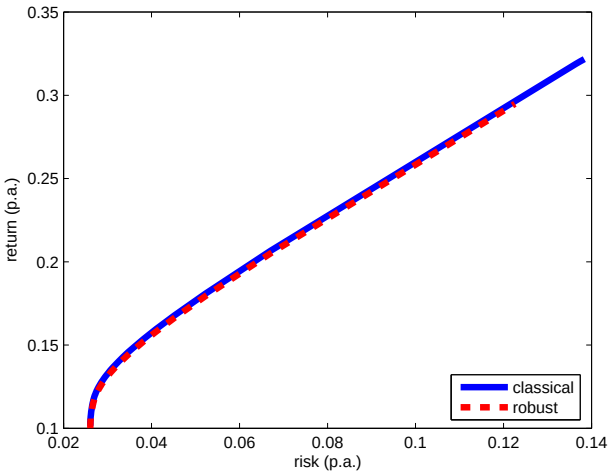


Figure 5.2: Illustration of the result of Proposition 5.7 that the robust efficient frontier is a shortened classical efficient frontier.

Proof. To prove that the classical and the robust efficient frontier coincide for the entire length of the robust efficient frontier, we show that the optimization problems to find the explicit points are equivalent.

To distinguish the two optimization problems with respect to the parameters tracing the efficient frontiers, we will use the parameter λ for the classical problem and θ for the robust one. Thus, we need to compare the following two problems:

$$\min_{x \in X} \quad (1-\lambda)\sqrt{x^T \hat{\Sigma} x} - \lambda(x^T \hat{\mu}) \tag{P_λ}$$

and

$$\min_{x \in X} \quad \left(1 - \theta + \theta \frac{\delta}{\sqrt{S}}\right) \sqrt{x^T \hat{\Sigma} x} - \theta(x^T \hat{\mu}). \tag{$RP_{\theta,\text{conf}}$}$$

As we want to find the tracing parameter for the classical optimal portfolio that corresponds to a given robust one, we fix the parameter θ. Defining

$$\lambda := \frac{\theta}{1 + \theta \frac{\delta}{\sqrt{S}}}$$

the classical problem

$$\min_{x \in X} \quad (1-\lambda)\sqrt{x^T \hat{\Sigma} x} - \lambda(x^T \hat{\mu})$$

reformulates to

$$\begin{aligned} &= \min_{x \in X} \quad \left(1 - \frac{\theta}{1 + \theta\frac{\delta}{\sqrt{S}}}\right) \sqrt{x^T \hat{\Sigma} x} - \frac{\theta}{1 + \theta\frac{\delta}{\sqrt{S}}} (x^T \hat{\mu}) \\ &= \min_{x \in X} \quad \frac{1}{1 + \theta\frac{\delta}{\sqrt{S}}} \left((1 - \theta + \theta\frac{\delta}{\sqrt{S}}) \sqrt{x^T \hat{\Sigma} x} - \theta (x^T \hat{\mu}) \right) \end{aligned}$$

which is equivalent to the robust formulation as the fraction $\frac{1}{1+\theta\frac{\delta}{\sqrt{S}}}$ is just a constant. □

Remark 5.8.

(i) Considering the special point $\theta = 0$, i.e. the (robust) minimum variance portfolio, we see that the corresponding λ is also zero. Both programs reduce to

$$\min_{x \in X} \quad \sqrt{x^T \hat{\Sigma} x}.$$

As the minimum variance portfolio does not depend on the uncertain parameter μ and as we did not explicitly consider uncertainty of the covariance matrix, the coincidence of the classical and the robust minimum variance portfolio was expected.

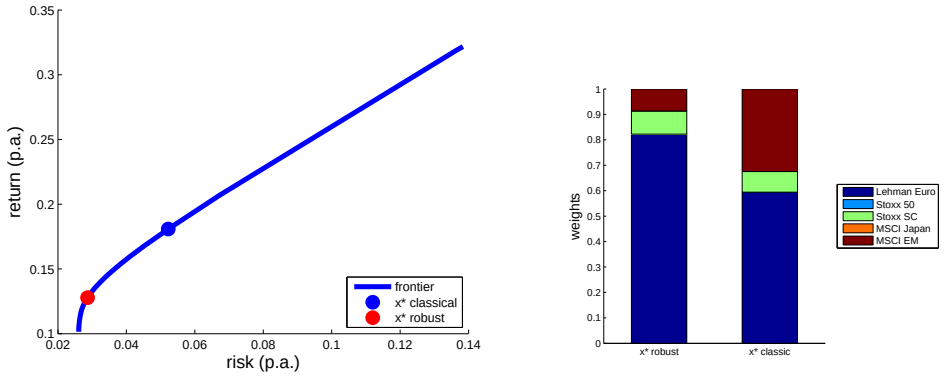
(ii) The result of Proposition 5.7 relies on the fact that the distribution (especially the second moment) of the maximum likelihood estimator $\hat{\mu}$ is again given in terms of $\hat{\Sigma}$, i.e. the matrices for measuring the portfolio risk and for describing the shape of the uncertainty set are the same. When applying the robust counterpart approach, we additionally obtain the expression $\lambda\delta\sqrt{x^T \hat{\Sigma} x}$ which can be interpreted as estimation risk and works as some kind of regularization. In Section 5.3 we will see an example where the matrix determining the uncertainty set is different from $\hat{\Sigma}$ and hence, the objective function cannot as nicely be combined.

The additional expression penalizes the investment in assets with large volatility, i.e. increases the influence of the risk of a portfolio compared to the expected return. Hence, such a robustification implies a shift of the trade-off between risk and return towards the less risky portfolios. This is illustrated in Figure 5.3 below.

(iii) Furthermore, the result holds as well if the risk of a portfolio is measured by the variance instead of the standard deviation, i.e. if the portfolio optimization problem is given in the form

$$\min_{x \in X} \quad (1 - \lambda) x^T \hat{\Sigma} x - \lambda (x^T \hat{\mu}).$$

Proposition 5.7 has the following impact for a particular investor: Let an investor have a certain risk aversion parameter λ which determines his personal trade-off between risk and return. By robustification of the portfolio optimization problem with a confidence ellipsoid, the investor changes his position on the efficient frontier – he becomes more conservative, i.e. risk averse, and hence his optimal robust portfolio moves towards the minimum variance portfolio. This phenomenon is illustrated in Figure 5.3, where Figure 5.3(a) demonstrates the respective positions on the efficient frontier and Figure 5.3(b) shows the optimal portfolio allocation corresponding to the particular trade-off or risk aversion parameter λ. As the robust optimal portfolio is a little closer to the MVP, the investment in the more secure bond index is slightly higher. For easier comparison with the portfolio allocations along the entire efficient frontier, we recall Figure 4.10 here, shown again in Figure 5.4.



(a) positions on efficient frontier (b) corresponding allocations

Figure 5.3: Implications of Proposition 5.7 for a particular investor.

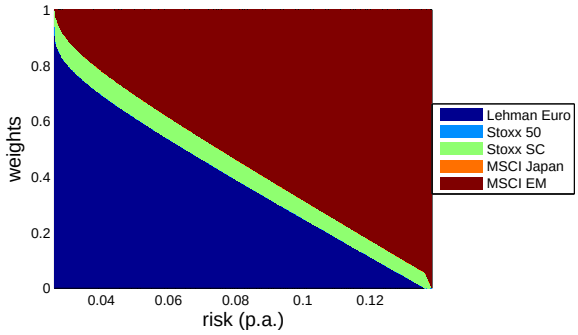


Figure 5.4: Portfolio allocations along the classical efficient frontier.

5.2.2 Joint confidence ellipsoid for μ and Σ

The naturally upcoming question is if this result of the robust efficient frontier being a shortened classical efficient frontier can be extended to the case of a joint uncertainty set for the mean vector and the covariance matrix. Before being able to prove this conjecture, we need to define a joint confidence ellipsoid around a point estimate for the pair (μ, Σ). To be able to create a confidence ellipsoid, we need to determine the distribution (more precisely, the first two moments in case of an elliptical distribution) of the point estimate $(\hat{\mu}, \hat{\Sigma})$. As for a normally distributed sample[1] the distribution of the (pair of) maximum likelihood estimators $(\hat{\mu}^{ML}, \hat{\Sigma}^{ML})$ is explicitly given, we consider in this subsection a sample $R_1, \ldots, R_S$ with $R_s \sim \mathcal{N}(\hat{\mu}, \hat{\Sigma})$ i.i.d.

Based on such a sample, the maximum likelihood estimators $\hat{\mu}^{ML}$ and $\hat{\Sigma}^{ML}$ are independent and have the following distributions, see Proposition 4.16:

$$\begin{aligned}\hat{\mu}^{ML} &\sim \mathcal{N}\left(\hat{\mu}, \frac{1}{S}\hat{\Sigma}\right),\\ \hat{\Sigma}^{ML} &\sim \mathcal{W}\left(\frac{1}{S}\hat{\Sigma}, S-1\right)\end{aligned}$$

with $\mathcal{W}(C, \nu)$ denoting the Wishart distribution with scale matrix C and ν degrees of freedom. For the definition and details about the Wishart distribution, see Appendix D.3. In Section 4.3 we have seen that the MLE for the covariance is a biased estimator whereas the sample estimator $\hat{\Sigma}^{SA} = \frac{S}{S-1}\hat{\Sigma}^{ML}$ is unbiased. Apart from the factor $\frac{S}{S-1}$, the sample and the maximum likelihood estimators for the covariance matrix are equal and for the mean vector they are identical anyway. As unbiasedness will greatly simplify the notation of the following calculations (by not having to explicitly carry through the multiplicative factor and its transformations) and hence improve readability, we will use the sample estimators instead of the MLEs. Thus, we start with having the following distribution of the independent sample estimators:

$$\begin{aligned}\hat{\mu}^{SA} &\sim \mathcal{N}\left(\hat{\mu}, \frac{1}{S}\hat{\Sigma}\right),\\ \hat{\Sigma}^{SA} &\sim \mathcal{W}\left(\frac{1}{S-1}\hat{\Sigma}, S-1\right).\end{aligned}$$

The moments of the Wishart distribution are given in Proposition D.6. To describe the covariance matrix of a matrix-valued random variable $A \in \mathbb{R}^{n\times n}$, the matrix A is transformed into an n^2-dimensional vector by stacking the columns of A successively underneath each other (see also Appendix C). Such a reformulated

[1] We assume that the result holds for general elliptical distributions, but we performed the explicit calculations only in case of a normal distribution.

vector will be denoted by $\text{vec}(A)$. Using this vector notation for a Wishart distributed random matrix, there exists a closed form expression for the covariance, see e.g. Meucci [57], page 85. Summarizing the moments of $\hat{\mu}^{SA}$ and $\text{vec}(\hat{\Sigma}^{SA})$, we hence get

$$\begin{aligned}
\mathbf{E}[\hat{\mu}^{SA}] &= \hat{\mu}, \\
\mathbf{Cov}[\hat{\mu}^{SA}] &= \frac{1}{S}\hat{\Sigma}, \\
\mathbf{E}[\text{vec}(\hat{\Sigma}^{SA})] &= (S-1)\cdot\frac{1}{S-1}\,\text{vec}(\hat{\Sigma}) = \text{vec}(\hat{\Sigma}), \\
\mathbf{Cov}[\text{vec}(\hat{\Sigma}^{SA})] &= (S-1)(I_{n^2}+K_{nn})\left(\frac{1}{S-1}\hat{\Sigma}\otimes\frac{1}{S-1}\hat{\Sigma}\right) \\
&= \frac{1}{S-1}(I_{n^2}+K_{nn})(\hat{\Sigma}\otimes\hat{\Sigma})
\end{aligned}$$

with K_{nn} denoting the commutation matrix and $\otimes$ representing the Kronecker product[2]. Note that the matrix $\mathbf{Cov}[\text{vec}(\hat{\Sigma}^{SA})]$ is not invertible (as the entire columns of $\hat{\Sigma}^{SA}$ are stacked underneath each other, all the off-diagonal elements appear twice in the vector and hence the covariance thereof must contain equal lines), but it is symmetric and positive semidefinite, i.e. a matrix decomposition $\mathbf{Cov}[\text{vec}(\hat{\Sigma}^{SA})] = MM$ can be found.

Proposition 5.9. *Let $\hat{\Sigma} \in S_+^n$.*

(i) The matrix $(I_{n^2}+K_{nn})(\hat{\Sigma}\otimes\hat{\Sigma})$ is symmetric.

(ii) Let $M := \frac{1}{\sqrt{2(S-1)}}(I_{n^2}+K_{nn})(\hat{\Sigma}^{\frac{1}{2}}\otimes\hat{\Sigma}^{\frac{1}{2}})$. Then it holds that

$$MM = \frac{1}{S-1}(I_{n^2}+K_{nn})(\hat{\Sigma}\otimes\hat{\Sigma}) = \mathbf{Cov}[\text{vec}(\hat{\Sigma}^{SA})].$$

Proof. In this proof we will need many of the properties and calculations for the Kronecker product which are summarized in Lemma C.7.

[2]In Appendix C the definitions thereof and some useful rules for calculation are summarized. Nevertheless, for better readability we shortly recall the according definitions here:
The commutation matrix K_{nk} is implicitly defined via the equation

$$\text{vec}(A) = K_{nk}\,\text{vec}(A^T)$$

for $A \in \mathbb{R}^{n\times k}$ and the Kronecker product for two arbitrary matrices $A \in \mathbb{R}^{n\times k}$ and $B \in \mathbb{R}^{p\times q}$ is given by the $np \times kq$ matrix

$$A\otimes B = \begin{pmatrix} A_{11}B & \cdots & A_{1k}B \\ \vdots & \ddots & \vdots \\ A_{n1}B & \cdots & A_{nk}B \end{pmatrix}.$$

(i) As I_{n^2}, K_{nn} and $\hat{\Sigma} \otimes \hat{\Sigma}$ are all symmetric themselves and it holds that

$$\begin{aligned}(I_{n^2} + K_{nn})(\hat{\Sigma} \otimes \hat{\Sigma}) &= I_{n^2}(\hat{\Sigma} \otimes \hat{\Sigma}) + K_{nn}(\hat{\Sigma} \otimes \hat{\Sigma}) \\ &= (\hat{\Sigma} \otimes \hat{\Sigma}) + K_{nn}(\hat{\Sigma} \otimes \hat{\Sigma})\end{aligned}$$

it suffices to show that $K_{nn}(\hat{\Sigma} \otimes \hat{\Sigma})$ is symmetric. Furthermore, as

$$[K_{nn}(\hat{\Sigma} \otimes \hat{\Sigma})]^T = (\hat{\Sigma} \otimes \hat{\Sigma})^T K_{nn}^T = (\hat{\Sigma} \otimes \hat{\Sigma}) K_{nn}$$

we hence need to prove that

$$K_{nn}(\hat{\Sigma} \otimes \hat{\Sigma}) = (\hat{\Sigma} \otimes \hat{\Sigma}) K_{nn}.$$

This is done if and only if

$$\operatorname{vec}(A)^T K_{nn}(\hat{\Sigma} \otimes \hat{\Sigma}) \operatorname{vec}(B) = \operatorname{vec}(A)^T (\hat{\Sigma} \otimes \hat{\Sigma}) K_{nn} \operatorname{vec}(B)$$

is satisfied for arbitrary matrices $A, B \in \mathbb{R}^{n \times n}$. Then, equality holds especially for matrices with a single entry of 1 and zeros otherwise, i.e. picking out individual entries of the matrix products in the middle.

Let $A, B \in \mathbb{R}^{n \times n}$ be arbitrary. We then get

$$\begin{aligned}\operatorname{vec}(A)^T K_{nn}(\hat{\Sigma} \otimes \hat{\Sigma}) \operatorname{vec}(B) &= \left[K_{nn}^T \operatorname{vec}(A)\right]^T (\hat{\Sigma} \otimes \hat{\Sigma}) \operatorname{vec}(B) \\ &\overset{(C.4)}{=} \left[\operatorname{vec}(A^T)\right]^T (\hat{\Sigma} \otimes \hat{\Sigma}) \operatorname{vec}(B) \\ &\overset{(C.11)}{=} \left[\operatorname{vec}(A^T)\right]^T \operatorname{vec}(\hat{\Sigma} B \hat{\Sigma}) \\ &\overset{(C.12)}{=} \operatorname{tr}(A \hat{\Sigma} B \hat{\Sigma})\end{aligned}$$

and

$$\begin{aligned}\operatorname{vec}(A)^T (\hat{\Sigma} \otimes \hat{\Sigma}) K_{nn} \operatorname{vec}(B) &\overset{(C.4)}{=} \operatorname{vec}(A)^T (\hat{\Sigma} \otimes \hat{\Sigma}) \operatorname{vec}(B^T) \\ &\overset{(C.11)}{=} \operatorname{vec}(A)^T \operatorname{vec}(\hat{\Sigma} B^T \hat{\Sigma}) \\ &\overset{(C.12)}{=} \operatorname{tr}(A^T \hat{\Sigma} B^T \hat{\Sigma}) \\ &= \operatorname{tr}\big([A^T \hat{\Sigma} B^T \hat{\Sigma}]^T\big) \\ &= \operatorname{tr}(\hat{\Sigma} B \hat{\Sigma} A) \\ &\overset{(C.1)}{=} \operatorname{tr}(A \hat{\Sigma} B \hat{\Sigma}),\end{aligned}$$

hence equality, and thus the matrix $(I_{n^2} + K_{nn})(\hat{\Sigma} \otimes \hat{\Sigma})$ is symmetric.

(ii) Let $M := \frac{1}{\sqrt{2(S-1)}}(I_{n^2} + K_{nn})(\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}})$. Exploiting symmetry of the matrices $(I_{n^2} + K_{nn})(\hat{\Sigma} \otimes \hat{\Sigma})$ and $(I_{n^2} + K_{nn})(\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}})$, it then holds that

$$\begin{aligned}
MM &= \frac{1}{2(S-1)}(I_{n^2} + K_{nn})(\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}}) \cdot (I_{n^2} + K_{nn})(\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}}) \\
&= \frac{1}{2(S-1)}(I_{n^2} + K_{nn})(\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}}) \cdot (\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}})(I_{n^2} + K_{nn}) \\
&\overset{(C.6)}{=} \frac{1}{2(S-1)}(I_{n^2} + K_{nn})(\hat{\Sigma} \otimes \hat{\Sigma})(I_{n^2} + K_{nn}) \\
&= \frac{1}{2(S-1)}(\hat{\Sigma} \otimes \hat{\Sigma})(I_{n^2} + K_{nn})(I_{n^2} + K_{nn}) \\
&= \frac{1}{2(S-1)}(\hat{\Sigma} \otimes \hat{\Sigma})(I_{n^2} + 2K_{nn} + \underbrace{K_{nn}^2}_{\overset{(C.5)}{=} I_{n^2}}) \\
&= \frac{1}{S-1}(\hat{\Sigma} \otimes \hat{\Sigma})(I_{n^2} + K_{nn}) \\
&= \frac{1}{S-1}(I_{n^2} + K_{nn})(\hat{\Sigma} \otimes \hat{\Sigma}) = \mathbf{Cov}[\mathrm{vec}(\hat{\Sigma}^{SA})]. \qquad \square
\end{aligned}$$

Having established the necessary prerequisites, we can create an ellipsoidal uncertainty set jointly for the mean vector and the covariance matrix. We first combine the two variables into one large vector by using the vector notation for the matrix $\hat{\Sigma}^{SA}$, i.e. the uncertain variable under consideration is expressed by

$$v^{SA} = \begin{pmatrix} \hat{\mu}^{SA} \\ \mathrm{vec}(\hat{\Sigma}^{SA}) \end{pmatrix} \in \mathbb{R}^{n+n^2}.$$

As the matrix $\mathbf{Cov}[\mathrm{vec}(\hat{\Sigma}^{SA})]$ is not invertible, we cannot use formula (5.1) as in the previous section to define the uncertainty set. But in Example 3.26 we have already seen that there is a different formulation for an ellipse which does not involve the inverse of the covariance matrix of $\mathrm{vec}(\hat{\Sigma}^{SA})$:

$$\mathcal{U}_\delta(\hat{v}) = \{v \in \mathbb{R}^{n+n^2} \mid v = \hat{v} + \delta V^{\frac{1}{2}} z, \|z\| \leq 1\} \tag{5.5}$$

with

$$\hat{v} = \mathbf{E}[v^{SA}] = \mathbf{E}\left[\begin{pmatrix} \hat{\mu}^{SA} \\ \hat{\Sigma}^{SA} \end{pmatrix}\right] = \begin{pmatrix} \hat{\mu} \\ \hat{\Sigma} \end{pmatrix} \tag{5.6}$$

describing the center of the uncertainty set, and V denoting the covariance matrix of v^{SA} which determines the shape of the ellipsoid. By exploiting knowledge about the distribution of v^{SA}, the above formulation of a joint confidence ellipsoid can be simplified. Especially independence of the two estimators $\hat{\mu}^{SA}$ and $\hat{\Sigma}^{SA}$ and

the explicit formulas for their individual covariance matrices are of use and allow a blockwise representation of the joint covariance V as

$$V = \begin{pmatrix} \frac{1}{S}\hat{\Sigma} & 0 \\ 0 & Q \end{pmatrix}$$

with

$$Q = \mathbf{Cov}(\mathrm{vec}(\hat{\Sigma}^{SA})) = \frac{1}{S-1}(I_{n^2} + K_{nn})(\hat{\Sigma} \otimes \hat{\Sigma}).$$

Analogous to the vector v being a composition of the vector μ and the matrix Σ, the auxiliary vector z is formed. Hence, we separate z into

$$z = \begin{pmatrix} z_1 \\ \mathrm{vec}(Z) \end{pmatrix}$$

with $z_1 \in \mathbb{R}^n$ and $Z \in S^n$, i.e. Z being a symmetric $n \times n$ matrix. Note that we can assume Z to be symmetric, as the uncertainty set around Σ is a subset of the space of symmetric and positive semidefinite matrices, otherwise the elements in $\mathcal{U}_\delta(\hat{v})$ cannot represent covariance matrices.

Based on these preliminary considerations the above uncertainty set $\mathcal{U}_\delta(\hat{v})$ can be rewritten in a more manageable form. This result is summarized in the following proposition.

Proposition 5.10. *Consider a joint uncertainty set for the pair (μ, Σ) (combined in a vector v) based on a confidence ellipsoid as given in Equation (5.5). This uncertainty set can be equivalently expressed as*

$$\begin{aligned}\mathcal{U}_\delta(\hat{\mu}, \hat{\Sigma}) = \{(\mu, \Sigma) \in \mathbb{R}^n \times S^n_+ \mid S(\mu - \hat{\mu})^T \hat{\Sigma}^{-1}(\mu - \hat{\mu}) + \\ \frac{S-1}{2} \|\hat{\Sigma}^{-\frac{1}{2}}(\Sigma - \hat{\Sigma})\hat{\Sigma}^{-\frac{1}{2}}\|^2_{tr} \leq \delta^2\}.\end{aligned}$$

Proof. According to Proposition 5.9 and the structure of V, the matrix $V^{\frac{1}{2}}$ is given by

$$V^{\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{S}}\hat{\Sigma}^{\frac{1}{2}} & 0 \\ 0 & \frac{1}{\sqrt{2(S-1)}}(I_{n^2} + K_{nn})(\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}}) \end{pmatrix}.$$

Using the decomposition of $z \in \mathbb{R}^{n+n^2}$ into the two parts

$$z = \begin{pmatrix} z_1 \\ \mathrm{vec}(Z) \end{pmatrix},$$

the equation

$$v = \hat{v} + \delta V^{\frac{1}{2}} z$$

can be expressed as

$$\begin{pmatrix} \mu \\ \operatorname{vec}(\Sigma) \end{pmatrix} = \begin{pmatrix} \hat{\mu} \\ \operatorname{vec}(\hat{\Sigma}) \end{pmatrix} + \delta \cdot \begin{pmatrix} \frac{1}{\sqrt{S}} \hat{\Sigma}^{\frac{1}{2}} z_1 \\ \frac{1}{\sqrt{2(S-1)}} (I_{n^2} + K_{nn})(\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}}) \operatorname{vec}(Z) \end{pmatrix}.$$

Thus, these are two equations that are coupled by the constraint

$$\|z\|_2^2 = \|z_1\|_2^2 + \|\operatorname{vec}(Z)\|_2^2 \overset{(C.2)}{=} \|z_1\|_2^2 + \|Z\|_{\text{tr}}^2 \leq 1.$$

Considering the upper equation, we can perform the same reformulations as in Example 3.26 and obtain

$$\begin{aligned} & \mu = \hat{\mu} + \delta \frac{1}{\sqrt{S}} \hat{\Sigma}^{\frac{1}{2}} z_1 \\ \Leftrightarrow \quad & (\mu - \hat{\mu})^T \hat{\Sigma}^{-1} (\mu - \hat{\mu}) = \frac{\delta^2}{S} \|z_1\|_2^2 \\ \Leftrightarrow \quad & \|z_1\|_2^2 = \frac{S}{\delta^2} (\mu - \hat{\mu})^T \hat{\Sigma}^{-1} (\mu - \hat{\mu}). \end{aligned}$$

The lower equation is given by

$$\operatorname{vec}(\Sigma) = \operatorname{vec}(\hat{\Sigma}) + \frac{\delta}{\sqrt{2(S-1)}} (I_{n^2} + K_{nn})(\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}}) \operatorname{vec}(Z),$$

or equivalently,

$$\operatorname{vec}(\Sigma - \hat{\Sigma}) = \frac{\delta}{\sqrt{2(S-1)}} (I_{n^2} + K_{nn})(\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}}) \operatorname{vec}(Z). \tag{5.7}$$

Having such a matrix (or vector) equality of the form $A = B$, it then also holds that $A^T C A = B^T C B$ for an arbitrary, suitably sized matrix C. Using this formulation with

$$C = \hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}$$

we obtain for the left hand side of Equation (5.7):

$$\begin{aligned} & \big(\operatorname{vec}(\Sigma - \hat{\Sigma})\big)^T (\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}) \operatorname{vec}(\Sigma - \hat{\Sigma}) \\ & \overset{(C.13)}{=} \big(\operatorname{vec}(\Sigma - \hat{\Sigma})\big)^T (I_n \otimes \hat{\Sigma}^{-1})(\hat{\Sigma}^{-1} \otimes I_n) \operatorname{vec}(\Sigma - \hat{\Sigma}) \\ & \overset{(C.8),(C.11)}{=} \big[\operatorname{vec}\big(\hat{\Sigma}^{-1}(\Sigma - \hat{\Sigma}) I_n\big)\big]^T \cdot \big[\operatorname{vec}\big(I_n (\Sigma - \hat{\Sigma}) \hat{\Sigma}^{-1}\big)\big] \\ & \overset{(C.12)}{=} \operatorname{tr}\Big(\big[\hat{\Sigma}^{-1}(\Sigma - \hat{\Sigma})\big]^T \cdot \big[(\Sigma - \hat{\Sigma}) \hat{\Sigma}^{-1}\big]\Big) \\ & = \operatorname{tr}\Big((\Sigma - \hat{\Sigma}) \hat{\Sigma}^{-1} (\Sigma - \hat{\Sigma}) \hat{\Sigma}^{-1}\Big) \\ & = \operatorname{tr}\Big((\Sigma - \hat{\Sigma}) \hat{\Sigma}^{-\frac{1}{2}} \hat{\Sigma}^{-\frac{1}{2}} (\Sigma - \hat{\Sigma}) \hat{\Sigma}^{-\frac{1}{2}} \hat{\Sigma}^{-\frac{1}{2}}\Big) \\ & = \operatorname{tr}\Big(\hat{\Sigma}^{-\frac{1}{2}} (\Sigma - \hat{\Sigma}) \hat{\Sigma}^{-\frac{1}{2}} \hat{\Sigma}^{-\frac{1}{2}} (\Sigma - \hat{\Sigma}) \hat{\Sigma}^{-\frac{1}{2}}\Big) \\ & = \|\hat{\Sigma}^{-\frac{1}{2}} (\Sigma - \hat{\Sigma}) \hat{\Sigma}^{-\frac{1}{2}}\|_{\text{tr}}^2. \end{aligned}$$

The analogous multiplications for the right hand side of Equation (5.7) can be simplified as follows using symmetry of $(I_{n^2} + K_{nn})(\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}})$:

$$
\begin{aligned}
&\frac{\delta}{\sqrt{2(S-1)}} \operatorname{vec}(Z)^T (\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}})(I_{n^2} + K_{nn}) \cdot (\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}) \\
&\qquad \cdot \frac{\delta}{\sqrt{2(S-1)}} (I_{n^2} + K_{nn})(\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}}) \operatorname{vec}(Z) \\
&\overset{(C.6)}{=} \frac{\delta^2}{2(S-1)} \operatorname{vec}(Z)^T (I_{n^2} + K_{nn})(\hat{\Sigma}^{\frac{1}{2}} \otimes \hat{\Sigma}^{\frac{1}{2}}) \\
&\qquad \cdot (\hat{\Sigma}^{-\frac{1}{2}} \otimes \hat{\Sigma}^{-\frac{1}{2}})(I_{n^2} + K_{nn}) \operatorname{vec}(Z) \\
&\overset{(C.6)}{=} \frac{\delta^2}{2(S-1)} \operatorname{vec}(Z)^T (I_{n^2} + K_{nn}) \underbrace{(I_n \otimes I_n)}_{I_{n^2}} (I_{n^2} + K_{nn}) \operatorname{vec}(Z) \\
&\overset{(C.4)}{=} \frac{\delta^2}{2(S-1)} \operatorname{vec}(Z)^T (I_{n^2} + K_{nn}) \big(\operatorname{vec}(Z) + \operatorname{vec}(Z^T)\big) \\
&\overset{(C.4)}{=} \frac{\delta^2}{2(S-1)} \operatorname{vec}(Z)^T \big(\operatorname{vec}(Z) + \operatorname{vec}(Z^T) + \operatorname{vec}(Z^T) + \operatorname{vec}(Z)\big) \\
&= \frac{\delta^2}{2(S-1)} \operatorname{vec}(Z)^T 4 \operatorname{vec}(Z) \\
&= \frac{2\delta^2}{S-1} \big[\operatorname{vec}(Z)^T \operatorname{vec}(Z)\big] \\
&\overset{(C.12)}{=} \frac{2\delta^2}{S-1} \cdot \operatorname{tr}(Z^T Z) \\
&= \frac{2\delta^2}{S-1} \, \|Z\|_{\mathrm{tr}}^2.
\end{aligned}
$$

Setting both sides equal yields

$$\|Z\|_{\mathrm{tr}}^2 = \frac{S-1}{2\delta^2} \, \|\hat{\Sigma}^{-\frac{1}{2}}(\Sigma - \hat{\Sigma})\hat{\Sigma}^{-\frac{1}{2}}\|_{\mathrm{tr}}^2.$$

Finally, using the coupling relation, we obtain

$$
\begin{aligned}
1 &\geq \|z_1\|_2^2 + \|Z\|_{\mathrm{tr}}^2 \\
&= \frac{S}{\delta^2}(\mu - \hat{\mu})^T \hat{\Sigma}^{-1}(\mu - \hat{\mu}) + \frac{S-1}{2\delta^2} \, \|\hat{\Sigma}^{-\frac{1}{2}}(\Sigma - \hat{\Sigma})\hat{\Sigma}^{-\frac{1}{2}}\|_{\mathrm{tr}}^2,
\end{aligned}
$$

or equivalently

$$S(\mu - \hat{\mu})^T \hat{\Sigma}^{-1}(\mu - \hat{\mu}) + \frac{S-1}{2} \|\hat{\Sigma}^{-\frac{1}{2}}(\Sigma - \hat{\Sigma})\hat{\Sigma}^{-\frac{1}{2}}\|_{\mathrm{tr}}^2 \leq \delta^2$$

which proves the statement. □

Having this particular uncertainty set $\mathcal{U}_\delta(\hat{\mu}, \hat{\Sigma})$, we now need to determine the worst case parameters thereof and use them to reformulate the robust optimization problem. We will find that as in the above case of having a confidence ellipsoid around the return vector only, the robust program reveals the same structure of the objective function as the classical problem, namely $f(x) = a\sqrt{x^T\hat{\Sigma}x} + bx^T\hat{\mu}$ with some factors a and b. Thus, an analogous proof as for Proposition 5.7 is applicable to show that also in this case of a joint confidence ellipsoid for (μ, Σ), the robust efficient frontier coincides with a part of the classical efficient frontier. The just verbally described calculations and proofs will be carried out explicitly in the subsequent propositions.

Proposition 5.11. *Let the joint uncertainty set for the parameters (μ, Σ) be given by a confidence ellipsoid around the MLEs $(\hat{\mu}, \hat{\Sigma})$ as described in Proposition 5.10. Then, the robust counterpart program to the portfolio optimization problem (P_λ) can be reformulated to*

$$\begin{aligned}
&\min_{x \in X} \quad \max_{(\mu,\Sigma) \in \mathcal{U}_\delta(\hat{\mu},\hat{\Sigma})} (1-\lambda)\sqrt{x^T\Sigma x} - \lambda x^T\mu \\
&= \min_{x \in X} \quad \max_{\kappa \in [0,1]} \left[(1-\lambda)\sqrt{1 + \delta\sqrt{\frac{2}{S-1}\,(1-\kappa)}} + \lambda\delta\frac{\sqrt{\kappa}}{\sqrt{S}} \right] \sqrt{x^T\hat{\Sigma}x} - \lambda x^T\hat{\mu}.
\end{aligned}$$

Proof. We first note that we can split up the joint uncertainty set

$$\mathcal{U}_\delta(\hat{\mu}, \hat{\Sigma}) = \left\{ S(\mu-\hat{\mu})^T\hat{\Sigma}^{-1}(\mu-\hat{\mu}) + \frac{S-1}{2}\,\|\hat{\Sigma}^{-\frac{1}{2}}(\Sigma-\hat{\Sigma})\hat{\Sigma}^{-\frac{1}{2}}\|_{\mathrm{tr}}^2 \le \delta^2 \right\}$$

into two separate ones with the respective sizes summing up to δ^2 again:

$$\begin{aligned}
\mathcal{U}(\hat{\mu}) &= \mathcal{U}_{\sqrt{\kappa}\delta}(\hat{\mu}) = \{\mu \in \mathbb{R}^n \mid S(\mu-\hat{\mu})^T\hat{\Sigma}^{-1}(\mu-\hat{\mu}) \le \kappa\delta^2\}, \\
\mathcal{U}(\hat{\Sigma}) &= \mathcal{U}_{\sqrt{1-\kappa}\delta}(\hat{\Sigma}) = \left\{ \Sigma \in S_+^n \mid \frac{S-1}{2}\,\|\hat{\Sigma}^{-\frac{1}{2}}(\Sigma-\hat{\Sigma})\hat{\Sigma}^{-\frac{1}{2}}\|_{\mathrm{tr}}^2 \le (1-\kappa)\delta^2 \right\}
\end{aligned}$$

with the additionally introduced variable $\kappa \in [0,1]$. With these definitions, the problem to determine the worst case parameters can be devided into several smaller optimization problems as follows:

$$\max_{(\mu,\Sigma) \in \mathcal{U}_\delta(\hat{\mu},\hat{\Sigma})} \quad (1-\lambda)\sqrt{x^T\Sigma x} - \lambda x^T\mu$$

is equivalent to

$$\max_{\kappa \in [0,1]} \quad \max_{\Sigma \in \mathcal{U}(\hat{\Sigma})} \quad \max_{\mu \in \mathcal{U}(\hat{\mu})} \quad (1-\lambda)\sqrt{x^T\Sigma x} - \lambda x^T\mu.$$

Note that for $\lambda = 0$ or $\lambda = 1$, the maximization over μ or Σ, respectively, is omitted. Hence, these two special cases are contained within the following general

calculations, and when solving the respective inner optimization problems over μ and Σ, we can without loss of generality assume that $\lambda > 0$ resp. $\lambda < 1$.

Now we solve two of the nested optimization problems successively, starting with the innermost one.

(i) Consider the optimization problem (which is only necessary in case $\lambda > 0$)

$$\max_{\mu \in \mathcal{U}(\hat{\mu})} \quad (1-\lambda)\sqrt{x^T \Sigma x} - \lambda x^T \mu. \tag{$P_{\text{aux},\mu}$}$$

From Example 3.26 we obtain that the optimal parameter μ^* is given by

$$\mu^* = \hat{\mu} - \delta \sqrt{\frac{\kappa}{S}} \frac{1}{\sqrt{x^T \hat{\Sigma} x}} \hat{\Sigma} x.$$

(ii) Next we want to solve the optimization problem

$$\max_{\Sigma \in \mathcal{U}(\hat{\Sigma})} \quad (1-\lambda)\sqrt{x^T \Sigma x} - \lambda x^T \mu^*.$$

which can equivalenty be formulated as

$$\begin{aligned} \max_{\Sigma \in S_+^n} \quad & (1-\lambda)\sqrt{x^T \Sigma x} - \lambda x^T \mu^* \\ \text{s.t.} \quad & \frac{S-1}{2} \, \|\hat{\Sigma}^{-\frac{1}{2}} (\Sigma - \hat{\Sigma}) \hat{\Sigma}^{-\frac{1}{2}}\|_{\text{tr}}^2 \leq (1-\kappa)\delta^2. \end{aligned}$$

Inserting the explicit expression for μ^* gives

$$\begin{aligned} \max_{\Sigma \in S_+^n} \quad & (1-\lambda)\sqrt{x^T \Sigma x} - \lambda x^T \hat{\mu} + \lambda \delta \sqrt{\frac{\kappa}{S}} \sqrt{x^T \hat{\Sigma} x} \\ \text{s.t.} \quad & \frac{S-1}{2} \, \|\hat{\Sigma}^{-\frac{1}{2}} (\Sigma - \hat{\Sigma}) \hat{\Sigma}^{-\frac{1}{2}}\|_{\text{tr}}^2 \leq (1-\kappa)\delta^2. \end{aligned} \tag{$P_{\text{aux},\Sigma}$}$$

Abbreviating the expressions in the objective that are independent of the variable Σ by $h(x,\kappa)$, i.e. $h(x,\kappa) = -\lambda x^T \hat{\mu} + \lambda\delta\sqrt{\kappa S^{-1}}\sqrt{x^T \hat{\Sigma} x}$, and defining $\widetilde{\Sigma} := \Sigma - \hat{\Sigma}$, the problem ($P_{\text{aux},\Sigma}$) is equivalent to

$$\begin{aligned} \max_{\widetilde{\Sigma} \in S_+^n} \quad & (1-\lambda)\sqrt{x^T \hat{\Sigma} x + x^T \widetilde{\Sigma} x} + h(x,\kappa) \\ \text{s.t.} \quad & \frac{S-1}{2} \, \|\hat{\Sigma}^{-\frac{1}{2}} \widetilde{\Sigma} \hat{\Sigma}^{-\frac{1}{2}}\|_{\text{tr}}^2 \leq (1-\kappa)\delta^2. \end{aligned} \tag{$P_{\text{aux},\Sigma}$-1}$$

Defining further $\bar{\Sigma} := \hat{\Sigma}^{-\frac{1}{2}} \widetilde{\Sigma} \hat{\Sigma}^{-\frac{1}{2}}$ the program changes to

$$\begin{aligned} \max_{\bar{\Sigma} \in S_+^n} \quad & (1-\lambda)\sqrt{x^T \hat{\Sigma} x + x^T \hat{\Sigma}^{\frac{1}{2}} \bar{\Sigma} \hat{\Sigma}^{\frac{1}{2}} x} + h(x,\kappa) \\ \text{s.t.} \quad & \|\bar{\Sigma}\|_{\text{tr}}^2 \leq \frac{2}{S-1} \, (1-\kappa)\delta^2. \end{aligned} \tag{$P_{\text{aux},\Sigma}$-2}$$

Obviously, as the square root function is monotonically increasing, the objective function of this problem ($P_{\text{aux},\Sigma}$-2) is maximized if and only if $x^T\hat{\Sigma}^{\frac{1}{2}}\bar{\Sigma}\hat{\Sigma}^{\frac{1}{2}}x$ is maximized. Thus, we let $y := \hat{\Sigma}^{\frac{1}{2}}x$ and solve the auxiliary problem

$$\begin{aligned} \max_{\bar{\Sigma}\in S_+^n} \quad & y^T\bar{\Sigma}y \\ \text{s.t.} \quad & \|\bar{\Sigma}\|_{\text{tr}}^2 \leq \frac{2}{S-1}\,(1-\kappa)\delta^2. \end{aligned} \tag{$P_{\text{aux},\Sigma}$-3}$$

We proceed analogous to Example 3.38 and obtain that the optimal parameter $\bar{\Sigma}^*$ for this auxiliary problem (and hence also for the problem ($P_{\text{aux},\Sigma}$-2)) is given by

$$\bar{\Sigma}^* = \delta\sqrt{\frac{2}{S-1}\,(1-\kappa)}\;\frac{y}{\|y\|}\cdot\frac{y^T}{\|y\|}.$$

Setting $z := \delta\sqrt{\frac{2}{S-1}\,(1-\kappa)}$ for ease of readability and incorporating the optimal solution $\bar{\Sigma}^*$ into program ($P_{\text{aux},\Sigma}$-2), this simplifies to

$$\begin{aligned} &\max_{\substack{\bar{\Sigma}\in S_+^n \\ \|\bar{\Sigma}\|_{\text{tr}}^2\leq z^2}} \quad (1-\lambda)\sqrt{x^T\hat{\Sigma}x + y^T\bar{\Sigma}y} + h(x,\kappa) \\ &= (1-\lambda)\sqrt{x^T\hat{\Sigma}x + zy^T\frac{y}{\|y\|}\cdot\frac{y^T}{\|y\|}y} + h(x,\kappa) \\ &= (1-\lambda)\sqrt{x^T\hat{\Sigma}x + z\|y\|^2} + h(x,\kappa) \\ &= (1-\lambda)\sqrt{(1+z)\,x^T\hat{\Sigma}x} + h(x,\kappa) \end{aligned}$$

where for the last equality it is used that

$$\|y\|^2 = y^Ty = x^T\hat{\Sigma}^{\frac{1}{2}}\hat{\Sigma}^{\frac{1}{2}}x = x^T\hat{\Sigma}x.$$

The proposition is finally proved by plugging back in all the definitions made along the way of the various calculations, i.e.

$$\min_{x\in X} \quad \max_{(\mu,\Sigma)\in\mathcal{U}_\delta(\hat{\mu},\hat{\Sigma})} (1-\lambda)\sqrt{x^T\Sigma x} - \lambda x^T\mu$$

reformulates to

$$\begin{aligned}
&\min_{x\in X} \quad \max_{\kappa\in[0,1]} \quad (1-\lambda)\sqrt{(1+z)\,x^T\hat{\Sigma}x} + h(x,\kappa) \\
&= \min_{x\in X} \quad \max_{\kappa\in[0,1]} \quad (1-\lambda)\sqrt{\left(1+\delta\sqrt{\frac{2}{S-1}\,(1-\kappa)}\right)x^T\hat{\Sigma}x} \\
&\qquad\qquad\qquad\qquad - \lambda x^T\hat{\mu} + \lambda\delta\sqrt{\frac{\kappa}{S}}\sqrt{x^T\hat{\Sigma}x} \\
&= \min_{x\in X} \quad \max_{\kappa\in[0,1]} \quad \left[(1-\lambda)\sqrt{1+\delta\sqrt{\frac{2}{S-1}\,(1-\kappa)}} + \lambda\delta\sqrt{\frac{\kappa}{S}}\right]\sqrt{x^T\hat{\Sigma}x} - \lambda x^T\hat{\mu},
\end{aligned}$$

which is the desired result. □

Notation 5.12. *For notational convenience we introduce the abbreviation*

$$K(\lambda) := \max_{\kappa\in[0,1]} \quad \left[(1-\lambda)\sqrt{1+\delta\sqrt{\frac{2}{S-1}\,(1-\kappa)}} + \lambda\delta\sqrt{\frac{\kappa}{S}}\right]. \tag{5.8}$$

Remark 5.13. *Note that it is not necessary for the subsequent Proposition 5.16 to explicitly calculate the maximizing κ. We only need that we obtain a unique κ for each fixed parameter λ of the robust optimization problem. This can e.g. be assured by strict concavity in κ.*

Lemma 5.14. *Let $\lambda \in [0,1]$ and let $\kappa^*(\lambda)$ denote the optimal solution of*

$$\max_{\kappa\in[0,1]} \quad (1-\lambda)\sqrt{1+\delta\sqrt{\frac{2}{S-1}\,(1-\kappa)}} + \lambda\delta\sqrt{\frac{\kappa}{S}}.$$

Then, $\kappa^*(\lambda)$ is unique for each $\lambda \in [0,1]$.

Proof. Let $\lambda \in [0,1]$ be arbitrary, but fixed. The (one-dimensional) optimization problem under consideration is described by

$$\max_{\kappa\in[0,1]} \quad f(\kappa)$$

with

$$f(\kappa) = (1-\lambda)\sqrt{1+\delta\sqrt{\frac{2}{S-1}\,(1-\kappa)}} + \lambda\delta\sqrt{\frac{\kappa}{S}}.$$

We now distinguish the three cases of $\lambda = 0$, $\lambda = 1$ and $0 < \lambda < 1$.

(i) Let $\lambda = 0$. Thus, the optimization problem reduces to

$$\max_{\kappa \in [0,1]} \quad \sqrt{1 + \delta \sqrt{\frac{2}{S-1}\,(1-\kappa)}}$$

which is obviously maximized for $\kappa^* = 0$, i.e. we obtain a unique solution. The optimal objective value is hence given by

$$K(0) = f(\kappa^*) = \sqrt{1 + \delta \sqrt{\frac{2}{S-1}}}.$$

(ii) Let $\lambda = 1$. Then $f(\kappa)$ is given by

$$f(\kappa) = \frac{\delta}{\sqrt{S}} \cdot \sqrt{\kappa}$$

which is uniquely maximized over $\kappa \in [0,1]$ for $\kappa^* = 1$. The expression $K(1)$ is thus given by

$$K(1) = \frac{\delta}{\sqrt{S}}.$$

(iii) Let $0 < \lambda < 1$. For $\kappa \in (0,1)$ the first derivative of $f(\kappa)$ becomes

$$\begin{aligned}
f'(\kappa) &= (1-\lambda) \cdot \frac{1}{2\sqrt{1 + \delta \sqrt{\frac{2}{S-1}\,(1-\kappa)}}} \cdot \frac{-\delta}{S-1} \cdot \frac{1}{\sqrt{\frac{2}{S-1}\,(1-\kappa)}} \\
&\quad + \frac{\lambda\delta}{\sqrt{S}} \cdot \frac{1}{2\sqrt{\kappa}} \\
&= -\underbrace{(1-\lambda) \cdot \frac{\delta}{2(S-1)}}_{=:c_1>0} \cdot \frac{1}{\sqrt{1 + \delta \sqrt{\frac{2}{S-1}\,(1-\kappa)}}\sqrt{\frac{2}{S-1}\,(1-\kappa)}} \\
&\quad + \underbrace{\frac{\lambda\delta}{2\sqrt{S}}}_{=:c_2>0} \cdot \frac{1}{\sqrt{\kappa}} \\
&= -c_1 \left(1 + \delta \sqrt{\frac{2}{S-1}\,(1-\kappa)}\right)^{-\frac{1}{2}} \left(\frac{2}{S-1}\,(1-\kappa)\right)^{-\frac{1}{2}} + c_2\,\frac{1}{\sqrt{\kappa}}
\end{aligned}$$

and thus we get for the second derivative

$$\begin{aligned} f''(\kappa) = -c_1 \Bigg[& \frac{1}{2} \underbrace{\left(1 + \delta \sqrt{\frac{2}{S-1}\,(1-\kappa)}\right)^{-\frac{3}{2}}}_{>0} \cdot \underbrace{\frac{\delta}{2(1-\kappa)}}_{>0} \\ & + \underbrace{\left(1 + \delta \sqrt{\frac{2}{S-1}\,(1-\kappa)}\right)^{-\frac{1}{2}}}_{>0} \cdot \underbrace{\left(\frac{2}{S-1}\,(1-\kappa)\right)^{-\frac{3}{2}}}_{>0} \cdot \frac{1}{S-1} \Bigg] \\ & - \frac{c_2}{2\kappa^{\frac{3}{2}}} \\ < 0. & \end{aligned}$$

Thus, $f(\kappa)$ is a strictly concave function on $(0,1)$ and for its derivatives at the boundaries $\kappa \to 0$ and $\kappa \to 1$ it holds

$$\begin{aligned} f'(\kappa) &\to +\infty && \text{for } \kappa \to 0, \\ f'(\kappa) &\to -\infty && \text{for } \kappa \to 1. \end{aligned}$$

Hence, $f(\kappa)$ has a unique optimal solution in $(0,1)$.

Altogether, we have that for each fixed λ we obtain a unique optimal $\kappa^*(\lambda)$. □

Before continuing, we analyze the function $K(\lambda)$ as introduced in Equation (5.8) some more.

Proposition 5.15. *Let $K(\lambda)$ with $\lambda \in [0,1]$ be given as in Equation (5.8). Then it holds:*

(i) The function $K : [0,1] \to \mathbb{R}_+$ is convex and continuous.

(ii) For $\delta \leq \sqrt{S}$, K is monotonically decreasing, and for $\delta > \sqrt{S}$, K possesses a minimum in $(0,1)$.

Proof. Recall that $K(\lambda)$ with $\lambda \in [0,1]$ is given by

$$K(\lambda) = \max_{\kappa \in [0,1]} \left[(1-\lambda) \sqrt{1 + \delta \sqrt{\frac{2}{S-1}\,(1-\kappa)}} + \lambda \delta \sqrt{\frac{\kappa}{S}} \right]$$

with the optimal solution $\kappa^*(\lambda)$ being unique for each $\lambda \in [0,1]$, see Lemma 5.14.

(i) The function $K(\lambda)$ is convex as the maximum of linear functions. Furthermore it is continuous according to Proposition 2.40.

(ii) According to Proposition 2.41, part (ii), the derivative of $K(\lambda)$ in case of a unique solution $\kappa^*(\lambda)$ is given by

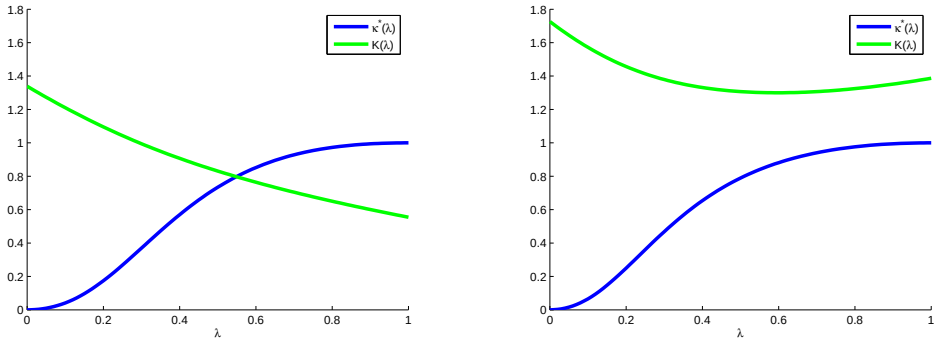
$$K'(\lambda) = -\sqrt{1 + \delta\sqrt{\frac{2}{S-1}\left(1-\kappa^*(\lambda)\right)}} + \delta\sqrt{\frac{\kappa^*(\lambda)}{S}}.$$

As for $\lambda = 0$ and $\lambda = 1$, the optimal solutions are $\kappa^*(0) = 0$ and $\kappa^*(1) = 1$, respectively, (see Lemma 5.14), the according derivatives of $K(\lambda)$ are given by

$$K'(0) = -\sqrt{1 + \delta\sqrt{\frac{2}{S-1}}},$$
$$K'(1) = -1 + \frac{\delta}{\sqrt{S}}.$$

If it holds that $\delta \leq \sqrt{S}$, both derivatives are negative (if $\delta = \sqrt{S}$, then $K'(1) = 0$) and together with convexity of $K(\lambda)$ we obtain that $K(\lambda)$ is monotonically descreasing. In case of $\delta > \sqrt{S}$, the derivative of K at $\lambda = 1$ is positive, hence, the minimum is attained in the open interval $(0, 1)$. □

Figure 5.5 illustrates the optimal solution $\kappa^*(\lambda)$ (obtained by numerical optimization) and the function $K(\lambda)$ for the two cases $\delta \leq \sqrt{S}$ and $\delta > \sqrt{S}$.



(a) case $\delta \leq \sqrt{S}$ (b) case $\delta > \sqrt{S}$

Figure 5.5: Illustration of $\kappa^*(\lambda)$ and $K(\lambda)$.

After having an explicit reformulation of the robust counterpart of the portfolio optimization problem when using a joint uncertainty set for (μ, Σ), we can show the same statement as in the case of an uncertainty set only around the return vector μ: the robust efficient frontier equals the classical efficient frontier, but is shortened.

Proposition 5.16. *Consider the portfolio optimization problem (P_λ) and let the joint uncertainty set for the parameters (μ, Σ) be given by a confidence ellipsoid around the MLEs as described in Proposition 5.10. Then for each $\theta \in [0,1]$ there exists a $\lambda \in \left[0, \frac{1}{1+\frac{\delta}{\sqrt{S}}}\right]$, $\lambda = \lambda(\theta) = \frac{\theta}{\theta + K(\theta)}$, such that the optimal solution $x^*_{rob}(\theta)$ of the corresponding local robust counterpart problem (see Proposition 5.11) equals the optimal solution $x^*_{cl}(\lambda)$ of the original problem.*

Proof. We proceed as in the proof of Proposition 5.7 by showing equivalence of the respective optimization problems. Recall the two optimization problems:

$$\min_{x \in X} \quad (1-\lambda)\sqrt{x^T \hat{\Sigma} x} - \lambda(x^T \hat{\mu}) \tag{P_λ}$$

and

$$\min_{x \in X} \quad \max_{\kappa \in [0,1]} \left[(1-\theta)\sqrt{1 + \delta\sqrt{\frac{2}{S-1}\,(1-\kappa)}} + \theta\delta\frac{\sqrt{\kappa}}{\sqrt{S}} \right] \sqrt{x^T \hat{\Sigma} x} - \theta x^T \hat{\mu}. \tag{$RP_{\theta,\text{conf}}$}$$

with θ being the tracing parameter in the robust program, analogous to the proof of Proposition 5.7.

To determine the parameter λ of the classical problem corresponding to a given tracing parameter θ of the robust program, we fix θ. For ease of notation and clarity of the proof, we again use the abbreviation

$$K(\theta) := (1-\theta)\sqrt{1 + \delta\sqrt{\frac{2}{S-1}\,(1-\kappa^*(\theta))}} + \theta\delta\,\frac{\sqrt{\kappa^*(\theta)}}{\sqrt{S}},$$

which is just a constant for fixed θ. With

$$\lambda = \lambda(\theta) := \frac{\theta}{\theta + K(\theta)}$$

the classical problem reformulates to

$$\begin{aligned}
&\min_{x \in X} \quad (1-\lambda)\sqrt{x^T \hat{\Sigma} x} - \lambda(x^T \hat{\mu}) \\
&= \min_{x \in X} \quad \left(1 - \frac{\theta}{\theta + K(\theta)}\right)\sqrt{x^T \hat{\Sigma} x} - \frac{\theta}{\theta + K(\theta)}(x^T \hat{\mu}) \\
&= \min_{x \in X} \quad \frac{1}{\theta + K(\theta)}\left(K(\theta)\sqrt{x^T \hat{\Sigma} x} - \theta(x^T \hat{\mu})\right)
\end{aligned}$$

which is equivalent to the robust formulation as $\frac{1}{\theta + K(\theta)}$ is merely a constant for fixed θ.

Finally, it remains to determine the maximum value for λ. For $\theta > 0$ we have

$$\lambda(\theta) = \frac{\theta}{\theta + K(\theta)} = \left(1 + \frac{K(\theta)}{\theta}\right)^{-1}$$

with the derivative

$$\lambda'(\theta) = -\underbrace{\left(1 + \frac{K(\theta)}{\theta}\right)^{-2}}_{>0} \cdot \frac{\theta K'(\theta) - K(\theta)}{\theta^2}.$$

This derivative is positive if and only if

$$\theta K'(\theta) - K(\theta) < 0$$

where $K'(\theta)$ is given explicitly in Proposition 5.15. To prove this inequality, note that since $\sqrt{\frac{2}{S-1}\,(1-\kappa^*(\theta))} \geq 0$ it holds that

$$0 < \sqrt{1 + \delta\sqrt{\frac{2}{S-1}\,(1-\kappa^*(\theta))}}. \tag{5.9}$$

Adding

$$\theta K'(\theta) = -\theta\sqrt{1 + \delta\sqrt{\frac{2}{S-1}\,\left(1-\kappa^*(\theta)\right)}} + \delta\theta\sqrt{\frac{\kappa^*(\theta)}{S}}$$

to both sides of inequality (5.9) yields

$$\begin{aligned}
\theta K'(\theta) &< \sqrt{1 + \delta\sqrt{\frac{2}{S-1}\,(1-\kappa^*(\theta))}} \\
&\quad - \theta\sqrt{1 + \delta\sqrt{\frac{2}{S-1}\,\left(1-\kappa^*(\theta)\right)}} + \delta\theta\sqrt{\frac{\kappa^*(\theta)}{S}} \\
&= (1-\theta)\sqrt{1 + \delta\sqrt{\frac{2}{S-1}\,(1-\kappa^*(\theta))}} + \delta\theta\sqrt{\frac{\kappa^*(\theta)}{S}} \\
&= K(\theta).
\end{aligned}$$

Thus, the derivative $\lambda'(\theta)$ is always positive, i.e. $\lambda(\theta)$ is a monotonically increasing function in θ and thus reaches is maximum value at $\theta = 1$. We can recall from Lemma 5.14 that $\kappa^*(1) = 1$, hence $K(1)$ simplifies to

$$K(1) = \delta\sqrt{\frac{\kappa^*(1)}{S}} = \frac{\delta}{\sqrt{S}}$$

which finally yields

$$\lambda_{\max} = \frac{1}{1 + K(1)} = \frac{1}{1 + \frac{\delta}{\sqrt{S}}}.$$

□

Note that even though an uncertainty set around $\hat{\Sigma}$ is incorporated, the minimum variance portfolio itself does not change. This holds since the robust optimization problem for $\lambda = 0$ reduces to

$$\min_{x \in X} \quad \sqrt{1 + \delta\sqrt{\frac{2}{S-1}}} \cdot \sqrt{x^T \hat{\Sigma} x}.$$

Consideration of uncertainty of the covariance matrix is reflected in the multiplicative factor $\sqrt{1 + \delta\sqrt{\frac{2}{S-1}}}$. Since this factor is larger than 1, we expect a higher volatility (i.e. objective value) of the minimum variance portfolio in the robust setting. This increase in the risk of the MVP can be interpreted as estimation risk. This fact is not illustrated in the figures, since for comparison with the classical efficient frontier, the point estimates $\hat{\mu}$ and $\hat{\Sigma}$ are used for plotting the robust efficient frontier as well. The robust return and covariance are merely necessary to determine the robust portfolio allocations.

Comparing the classical efficient frontier with the two robust frontiers obtained using confidence ellipsoids around $\hat{\mu}$ and $(\hat{\mu}, \hat{\Sigma})$, respectively, we observe the following:

- The minimum variance portfolio is the same in all three cases.
- Both robust efficient frontiers coincide with the classical efficient frontier, up to the trade-off parameter $\lambda = \frac{1}{1+\frac{\delta}{\sqrt{S}}}$ of the classical frontier. Note that for both trade-off parameters $\theta_\mu = 1$ and $\theta_{\mu,\Sigma} = 1$ at the right end of the two robust efficient frontiers the same maximum value for λ is attained:

  (i) When robustifying only the return parameter μ, the mapping from the (robust) trade-off parameter θ_μ to the (classical) trade-off parameter λ is given by (see Proposition 5.7)

  $$\lambda(\theta_\mu) = \frac{\theta_\mu}{1 + \theta_\mu \frac{\delta}{\sqrt{S}}}$$

  and hence $\lambda(1) = \frac{1}{1+\frac{\delta}{\sqrt{S}}}$.

  (ii) In case of robustifying jointly (μ, Σ), the mapping from $\theta_{\mu,\Sigma}$ to λ (see Proposition 5.16) is described by

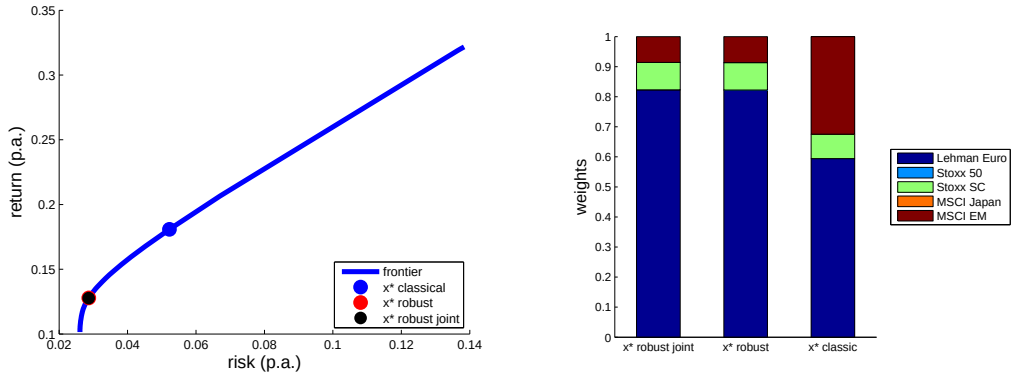
  $$\lambda(\theta_{\mu,\Sigma}) = \frac{\theta_{\mu,\Sigma}}{\theta_{\mu,\Sigma} + K(\theta_{\mu,\Sigma})}$$

  and since $K(1) = \frac{\delta}{\sqrt{S}}$ (Lemma 5.14), we obtain $\lambda(1) = \frac{1}{1+\frac{\delta}{\sqrt{S}}}$ as well.

- Between the left and the right end point, the two robust efficient frontiers map the trade-off parameter differently. This will be investigated more closely in the following.

- In case of robustification around μ only, the size δ of the uncertainty set reflects the probability of the parameter lying within the ellipse, i.e. the confidence. When having a joint uncertainty set for μ and Σ, the size δ does not correspond to the same confidence as before, i.e. it has to be interpreted differently with respect to representing probabilities. In the following, we compare uncertainty sets having the same value of δ.

As before, we consider an investor having a particular trade-off parameter λ. Using confidence ellipsoids around $\hat{\mu}$ and $(\hat{\mu}, \hat{\Sigma})$, respectively, we obtain the situation shown in Figure 5.6.



(a) positions on efficient frontier (b) corresponding allocations

Figure 5.6: Implications of Propositions 5.7, 5.16 for a particular investor.

The fact that the portfolio obtained by robustification of μ and Σ is always left of the portfolio obtained by robustification of the parameter μ only is hardly recognizable in the figure, but will be proved in the following proposition.

Proposition 5.17. *Let a trade-off parameter $\theta \in [0,1]$ be given. Furthermore, let $\lambda_1(\theta) = \frac{\theta}{1+\theta\frac{\delta}{\sqrt{S}}}$ denote the corresponding robust trade-off parameter in the setting of Proposition 5.7 with an uncertainty set of size δ for μ only, and let $\lambda_2(\theta) = \frac{\theta}{\theta+K(\theta)}$ be the robust trade-off parameter in the setting of Proposition 5.16 with a joint uncertainty set of the same size δ. Then it holds that $\lambda_2(\theta) \leq \lambda_1(\theta)$.*

Proof. First recall that since the robust optimal portfolios in both settings are lying on the classical efficient frontier, it suffices to relate the respective positions which are expressed in terms of the trade-off parameters $\lambda_1(\theta)$ and $\lambda_2(\theta)$.

Given a certain risk aversion (i.e. trade-off parameter) θ, Propositions 5.7 and 5.16 state the explicit formulas for determining the new position on the classical efficient frontier when a robust portfolio optimization is performed:

$$\lambda_1(\theta) = \frac{\theta}{1 + \theta\frac{\delta}{\sqrt{S}}}$$
$$\lambda_2(\theta) = \frac{\theta}{\theta + K(\theta)}$$

with $K(\theta)$ as given in Equation (5.8).

To prove that $\lambda_2(\theta) \leq \lambda_1(\theta)$ for all $\theta \in [0,1]$, it suffices to compare the denominators and hence show that

$$\theta + K(\theta) \geq 1 + \theta\frac{\delta}{\sqrt{S}},$$

or equivalently that

$$H(\theta) := \theta + K(\theta) - 1 - \theta\frac{\delta}{\sqrt{S}} \geq 0, \quad \forall\theta \in [0,1].$$

Using result from Lemma 5.14, we already know the following about the function $H(\theta)$:

$$H(0) = K(0) - 1 = \sqrt{1 + \delta\sqrt{\frac{2}{S-1}}} - 1 > 0$$
$$H(1) = K(1) - \frac{\delta}{\sqrt{S}} = 0.$$

Furthermore, $H(\theta)$ is convex (recall that $K(\theta)$ is convex according to Proposition 5.15). For the derivative of $H(\theta)$, we obtain that

$$\begin{aligned} H'(\theta) &= 1 + K'(\theta) - \frac{\delta}{\sqrt{S}} \\ &= 1 - \sqrt{1 + \delta\sqrt{\frac{2}{S-1}\left(1 - \kappa^*(\theta)\right)}} + \delta\sqrt{\frac{\kappa^*(\theta)}{S}} - \frac{\delta}{\sqrt{S}} \\ &= \underbrace{1 - \sqrt{1 + \delta\sqrt{\frac{2}{S-1}\left(1 - \kappa^*(\theta)\right)}}}_{<0} + \frac{\delta}{\sqrt{S}}\underbrace{\left(\sqrt{\kappa^*(\theta)} - 1\right)}_{<0} \\ &< 0. \end{aligned}$$

Thus, $H(\theta)$ is monotonically decreasing and convex on $[0,1]$ with $H(0) > 0$ and $H(1) = 0$ and hence does not possess a minimum on $(0,1)$. Therefore, $H(\theta) \geq 0$ for all $\theta \in [0,1]$ which gives the desired result of $\lambda_2(\theta) \leq \lambda_1(\theta)$. □

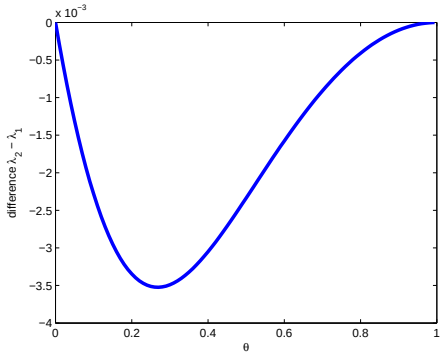


Figure 5.7: Difference between the positions of the robust portfolios.

Figure 5.7 illustrates the difference $\lambda_2(\theta) - \lambda_1(\theta)$ between the positions of the robust portfolios on the efficient frontier.

To conclude this section we shortly summarize the results. We have defined uncertainty sets via confidence ellipsoids, first only for the uncertain vector μ and then as well for the pair of uncertain parameters (μ, Σ). In both cases we found that the robust efficient frontier coincides with the classical efficient frontier up to some risk level, and we furthermore showed that robustifying both parameters leads to a more conservative portfolio allocation than robustification of μ only. This nice structural result that the robust and the classical efficient frontier are the same leads to the conclusion that the classical efficient frontier itself already consists of robust portfolio allocations. The proofs of these statements rely on the fact that the confidence ellipsoids are formed using the same matrix structure that is used to measure the portfolio's risk.

5.3 Combination of various statistical estimators

A completely different approach to create an uncertainty set for the return vector μ is to make use of several statistical estimators for the parameter. It is not clear in case of general elliptical distributions, why for example the maximum likelihood estimator should be preferred to any of the other estimators presented in Section 4.3. Thus, we want to take them equally into account and create an ellipsoidal uncertainty set such that all considered estimators are lying within. Let the set M denote the set of different estimates for the parameter μ, i.e.

$$M := \{\hat{\mu}^{ML}, \hat{\mu}^{ME}, \hat{\mu}^{QR}, \hat{\mu}^{TM}, \hat{\mu}^{HU}\}.$$

The first intuitive idea to create an uncertainty set containing the points in M would be to use the convex hull and define

$$\mathcal{U} = \mathrm{conv}\big(\hat{\mu}^{ML}, \hat{\mu}^{ME}, \hat{\mu}^{QR}, \hat{\mu}^{TM}, \hat{\mu}^{HU}\big).$$

Knowing that ellipsoidal uncertainty sets are more promising than polyhedral ones (Theorem 3.28), we therefore create an ellipsoid containing the estimates instead of using the convex hull. Thus, we consider the following uncertainty set:

$$\mathcal{U}_{\text{est}} = \{\mu \mid (\mu - \bar{\mu})^T \bar{\Sigma}^{-1} (\mu - \bar{\mu}) \leq \bar{\delta}^2\}$$

$$\text{with} \quad \bar{\mu} = \frac{1}{|M|} \sum_{m \in M} m$$

$$\bar{\Sigma} = \text{diag}(\bar{\sigma}_{11}^2, \ldots, \bar{\sigma}_{nn}^2) \quad \text{where} \quad \bar{\sigma}_{ii}^2 = \frac{1}{|M| - 1} \sum_{m \in M} (m_i - \bar{\mu}_i)^2$$

$$\bar{\delta}^2 = \max_{m \in M} (m - \bar{\mu})^T \bar{\Sigma}^{-1} (m - \bar{\mu}).$$

The following Figure 5.8 illustrates such an uncertainty set in the case of the two assets bond and stock. As the shape matrix $\bar{\Sigma}$ is given by a diagonal matrix, it is obvious that the axes of the ellipse are parallel to the coordinate axes. Furthermore, the lower (resp. higher) volatility of the bond (resp. stock) market is reflected by the shape of the ellipse, i.e. by the lengths of its axes.

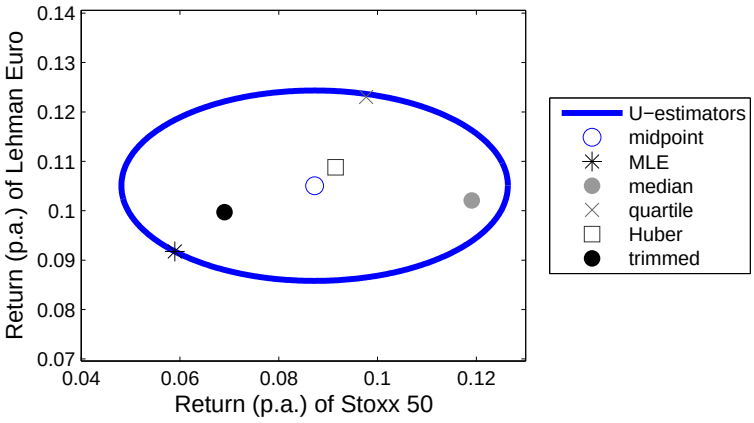


Figure 5.8: Illustration of the uncertainty set created by using different statistical estimators.

As this uncertainty set $\mathcal{U}_{\text{est}}$ naturally has the same structure as the confidence ellipsoid (just with different midpoint and shape matrix), the worst case parameter μ_{wc} can be deduced by analogous calculations and is given by

$$\mu_{wc} = \bar{\mu} - \bar{\delta} \frac{1}{\sqrt{x^T \bar{\Sigma} x}} \bar{\Sigma} x. \tag{5.10}$$

Thus, the robust optimization problem reformulates to

$$\begin{aligned} &\min_{x\in X} \quad \max_{\mu\in\mathcal{U}_{\text{est}}} (1-\lambda)\sqrt{x^T\hat{\Sigma}x} - \lambda x^T\mu \\ =&\min_{x\in X} \quad (1-\lambda)\sqrt{x^T\hat{\Sigma}x} - \lambda x^T\bar{\mu} + \bar{\delta}\lambda\sqrt{x^T\bar{\Sigma}x}. \end{aligned}$$

Compared to the classical problem the additional expression $\bar{\delta}\lambda\sqrt{x^T\bar{\Sigma}x}$ can be interpreted as a penalty term. As the matrix $\bar{\Sigma}$ contains the variances of the different estimators for each asset, the robust optimization problem penalizes investment in assets where the considered statistical estimators yield rather different values, i.e. lie further apart from each other.

Remark 5.18. *Note that for this uncertainty set, experts' opinions about point estimates for the return vector could easily be incorporated by treating them like additional estimators.*

Remark 5.19. *The uncertainty set $\mathcal{U}_{est}$ can become more sophisticated by additionally allowing the ellipsoid to be rotated, i.e. it does not necessarily lie parallel to the coordinate axes anymore. To find the smallest rotated ellipsoid containing all the estimators, an additional optimization problem has to be solved.*

A further alternative to create an uncertainty set containing a prescribed number of given points is to solve the optimization problem for a minimum volume ellipsoid. This results in a semidefinite program (SDP). Note that this problem has only a non-degenerate solution if the number of (independent) point estimates exceeds the number of assets. In case there are not sufficiently many estimators for the return vector, additional constraints have to be artificially introduced, like e.g. the length of each axis of the ellipsoid has to be strictly positive.

For a presentation and comparison of these alternative approaches see the diploma thesis of Middelkamp [61].

In the following the effect of robustification using such an uncertainty set for the return vector is illustrated. To compare with the results presented in Section 4.4, we again consider the market at the time point 01.11.2003 and we restrict the feasibility set to be given by

$$X = \{x \in \mathbb{R}^n \mid x^T\mathbf{1} = 1, x \geq 0\}.$$

The robust efficient frontier is plotted using the robust allocations and the maximum likelihood estimators to calculate the respective risk and return characteristics. (This means the expected portfolio return is plotted, not the expected robust portfolio return which is expressed in the optimization problem; the robust formulation was only needed to determine the weights.) Figure 5.9 shows both the classical and the robust efficient frontier, and in Figure 5.10 the associated allocations are plotted.

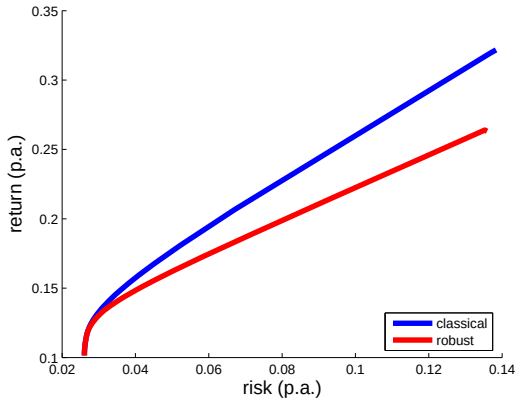


Figure 5.9: Classical and robust efficient frontier on 01.11.2003.

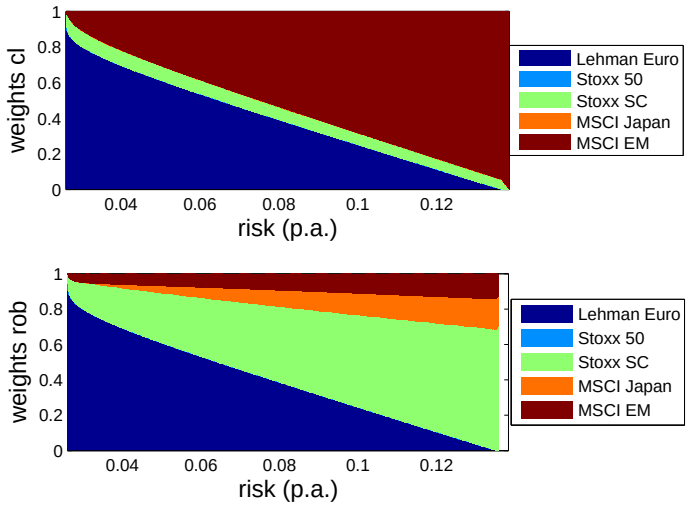
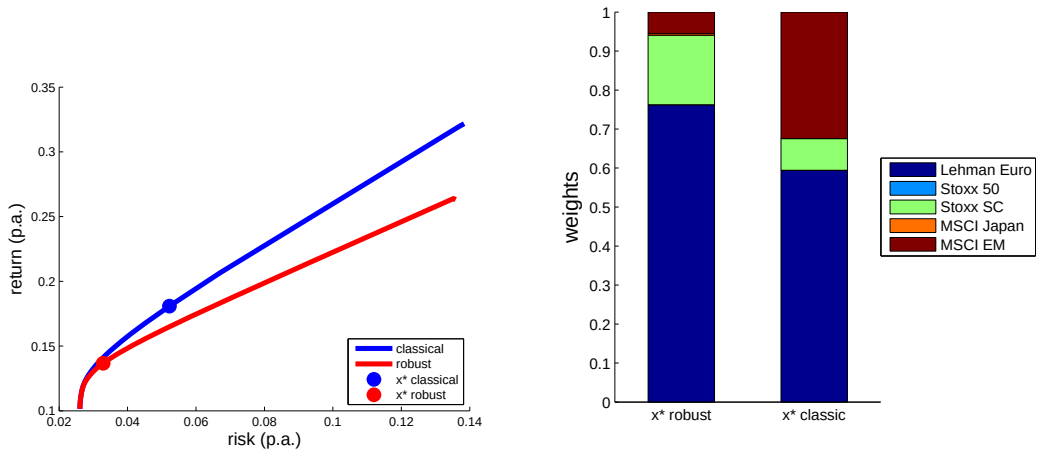


Figure 5.10: Classical and robust efficient portfolios on 01.11.2003.

Note that when defining an uncertainty set based on estimators, we obtain a rather different efficient frontier in the robust case and not – as in the previous section when using a confidence ellipsoid – the same as the classical one. Since the robust efficient frontier is a little shorter than the classical one, the remaining part at the right of the weight plot in the robust case is empty. It can also be seen that the robust portfolios are more diversified since they are invested in four of the five assets for the most part of the efficient frontier. The classical portfolios have an investment of roughly 10% in the Stoxx small caps, and the remaining 90% are moved from the most secure bond index to the riskier emerging market. In the robust allocations the emerging market never even reaches a portion of 20%.

Also for this type of robustificaton, we can analyze the influence on the optimal portfolio of a particular investor. Analogous to before we plot in Figure 5.11 the changed position on the robust efficient frontier and the according optimal allocation in comparison to the classical one. Analogous to the case of using confidence ellipsoids as uncertainty sets, the investor chooses are more conservative portfolio when performing a robust optimization. This general fact is observable in all considered robustifications, independent of the particular specification of the employed uncertainty set. In Chapter 7 we will investigate approaches combining market and expert information to obtain uncertainty sets, and in those cases we will also find that robustification leads to portfolios lying closer towards the minimum variance portfolio.



(a) positions on efficient frontiers (b) corresponding allocations

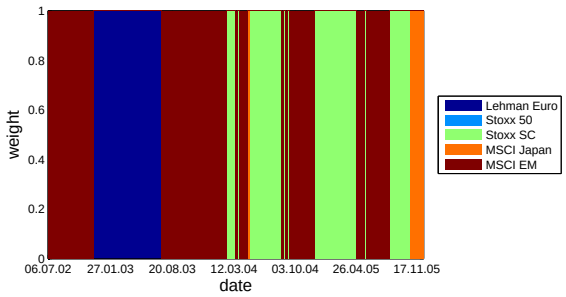
Figure 5.11: Implications of robust optimization for a particular investor.

Similar to the illustrations in Section 4.4 we not only want to investigate the portfolio allocations along the efficient frontier for a fixed point in time, but also the behavior of selected individual portfolios over time. Particular portfolios are naturally both ends of the efficient frontier, i.e. the minimum variance portfolio (MVP) and the maximum return portfolio (MRP). We additionally include the maximum sharpe ratio portfolio (MSRP) in the presentation.

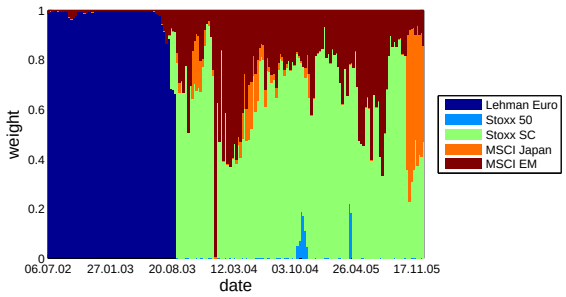
First of all it is worth noting that the classical and the robust minimum variance portfolios are identical for each point in time. This is obvious as the MVP is independent of the estimator for the return or an uncertainty set thereof, and because we did not impose any explicit uncertainty about the covariance matrix.

Naturally, robustification (around the vector of expected returns) is having the largest effect on the maximum return portfolio. In Figure 5.12 it can be seen that the classical and the robust maximum return portfolios differ substantially, except in the market phase where both approaches are fully invested in the bond

index, i.e. the classical MRP acts as conservatively as the robust MRP. Usually the robust maximum return portfolio is not as extreme as the classical MRP which consists only of the asset with the highest expected return. The robust approach is more defensely and diversifies the investment. Only in cases where there is a large distance between the highest and the second highest return – large enough such that even if an uncertainty set is put around the highest component, the second highest is still not contained within – the robust MRP consists of only one asset as well. Figure 5.12(a) illustrates again the weights of the classical MRP for easier comparison, and in plot 5.12(b) the allocation of the robust MRP over time is graphed.



(a) classical MRP

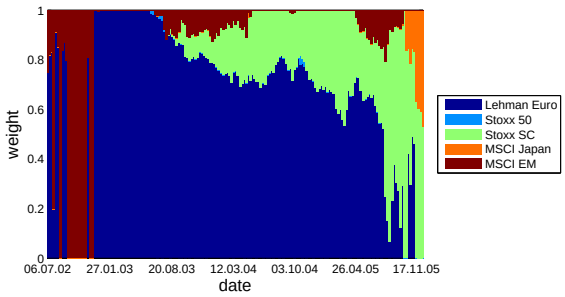


(b) robust MRP

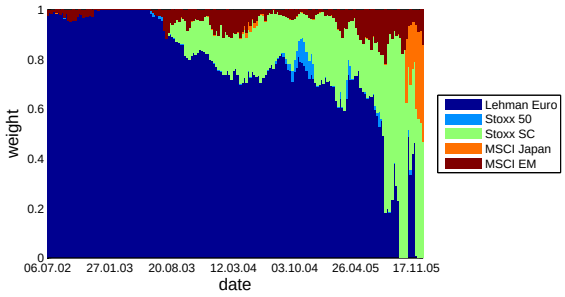
Figure 5.12: Classical and the robust MRP allocations over time.

Analogously we can compare a portfolio somewhere between the MVP and the MRP, the maximum sharpe ratio portfolio. Figure 5.13 shows again both the classical and the robust portfolios. They are rather similar most of the time, except at the beginning where the classical allocation changes its proportional investment in the emerging market rather often. Such an alternating behavior

suggests approximately equal performance of the two assets. In these cases where estimation errors can have a large effect on the optimal solution, the robust allocation is much smoother than the classical one.



(a) classical MSRP



(b) robust MSRP

Figure 5.13: Classical and the robust MSRP allocations over time.

Additional to merely comparing the portfolio weights over time, we also analyze the out-of-sample performance of the selected portfolios – illustrated only in case of the maximum return portfolio as there the difference between the classical and the robust allocation is the largest. For the maximum sharpe ratio the effects are similar to those occuring in case of the MRP, but milder.

Figure 5.14 shows the cumulated out-of-sample performance based on the historical data set. This means, at each point in time, the last 52 data points were used to calculate the parameters $\hat{\mu}$ and $\hat{\Sigma}$ or the uncertainty set, upon which the optimal allocations were deduced. These classical and robust portfolios were held for one period (i.e. one week) and the actually achieved returns were calculated using the current asset returns for this period.

It can be observed that in cases of bear markets the robust MRP acts a lot

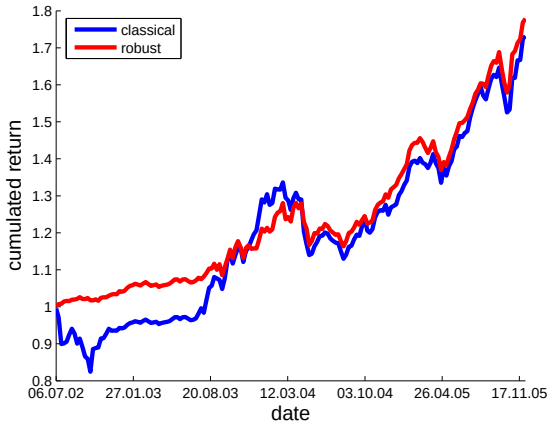


Figure 5.14: Out-of-sample performance of the classical and the robust MRP.

more conservative than the classical one – as already discussed above. As it is only invested in the bond at the beginning, it steadily makes the bond's small return and does not suffer the losses of the stock indices. The classical MRP is in some parts invested in the emerging market index and thus also takes those losses. In case of a bull market – towards the end of the time period – the classical MRP can participate more in the substantial gains of individual assets as it is always invested fully in the best one; best with respect to the previous year's time, but if this asset still performs highest in the following period, the classical MRP has exactly that return as well. The robust MRP hardly ever invests only in one asset, hence it usually does not participate with the full amount in the best asset's performance. Nevertheless, as the classical MRP loses almost 20% at the beginning and again around April 2004, the robust maximum return portfolio has a larger cumulative return throughout most of the time.

Furthermore, Table 5.1 summarizes the annualized in-sample and out-of-sample average characteristics for the minimum variance portfolio (whose allocation is identical in the classical and the robust setting), the maximum sharpe ratio portfolio and the maximum return portfolio. The displayed values are rounded to one digit which might lead to presumingly identical values even if there are differences. This happens for example in case of the MVP where the in-sample and out-of-sample Sharpe ratios seem to be the same even though the volatilities are different. The in-sample Sharpe ratio is actually 1.74 whereas the out-of-sample Sharpe ratio is 1.69.

It can be observed that both the classical and the robust portfolio mostly promise "better" figures in-sample than they can achieve out-of-sample, i.e. they expect higher returns and lower volatility in-sample than are actually realized out-of-sample. The exception here in this historical sample is the minimum variance portfolio which meets its expectation rather well. However, the discrepancy

		in-sample		out-of-sample	
		classical	robust	classical	robust
MVP	return	6.5%	6.5%	6.5%	6.5%
	volatility	2.4%	2.4%	2.7%	2.7%
	Sharpe ratio	1.7	1.7	1.7	1.7
MSP	return	11.9%	11.0%	10.5%	10.0%
	volatility	4.6%	3.3%	7.0%	4.5%
	Sharpe ratio	2.6	2.5	1.2	1.8
MRP	return	22.0%	19.5%	16.9%	17.2%
	volatility	11.5%	8.0%	13.4%	9.1%
	Sharpe ratio	1.9	1.9	1.1	1.7

Table 5.1: Averaged annualized in-sample and out-of-sample characteristics.

between the in-sample and out-of-sample numbers is smaller in case of the robust portfolio, as can be compared best by the values of the respective Sharpe ratios, hence the robust portfolios seem to be more trustworthy with respect to their expected characterizations.

To quantify the necessary allocation changes over time for a fixed portfolio, we determine and plot the *turnover* of the classical and the robust maximum return portfolio. The turnover is calculated along the time axis as the cumulated sum of the (absolute) weight changes from one time point to the next, i.e. as the sum of

$$\text{turnover}_t = \frac{1}{2}\|x_t^*(\lambda) - x_{t-1}^*(\lambda)\|_1$$

where $x_t^*(\lambda)$ denotes the optimal portfolio at time t to the parameter λ. The factor $\frac{1}{2}$ is included as a normalization such that completely selling and afterwards buying the entire portfolio results in a turnover of 1. Note that for the classical maximum return portfolio the turnover at each point in time is either zero or one, as either the same asset yields the highest return and thus the portfolio does not change, or else the entire investment in the previously best asset is sold and a new asset is bought. This fact can also be seen in Figure 5.15, as the curve for the turnover of the classical portfolio is a step function. In case of the robust portfolio, the turnover changes more smoothly, since the robust MRP is mostly not invested in merely one asset and hence smaller amounts of individual assets are modified.

Generally, Figure 5.15 expresses that over time a lot less allocation changes are needed in the robust MRP compared to the classical MRP. This is a desirable fact, as the turnover can be a measure for transaction costs. Figure 5.16 illustrates the out-of-sample performance including transaction costs that were approximated using the turnover. The cost factor for a complete turnover of the portfolio was set to 2%, representing a situation where no fixed transaction costs apply and the

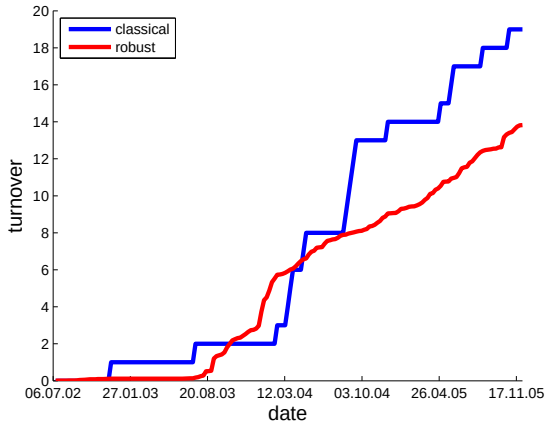


Figure 5.15: Turnover of the classical and the robust MRP.

same variable cost is assumed for all assets both for buying and selling. Hence, the change in allocation at each point in time multiplied by the cost factor measures the losses in the overall performance. This is reflected in Figure 5.16 where both the classical and the robust portfolio have a lower cumulated performance compared to the results without transaction costs shown in Figure 5.14. But as the turnover of the robust MRP is significantly smaller than the turnover of the classical MRP, the respective costs are more limited which results in a smaller performance reduction compared to the classical case. Hence, the advantage of the robust approach is even more evident.

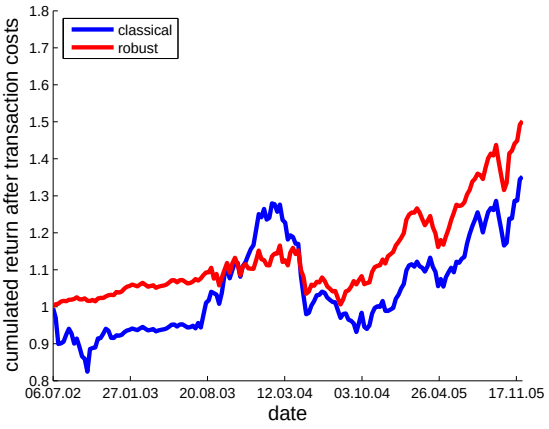


Figure 5.16: Out-of-sample performance of the classical and the robust MRP including approximate transaction costs.

A further analysis concerns the size of the uncertainty set which represents the conservativeness of the robustification. Figure 5.17 illustrates the size δ over

time. The value of δ is mostly around 3 but changes depending on the reliability of the parameter estimates on the underlying data sample.

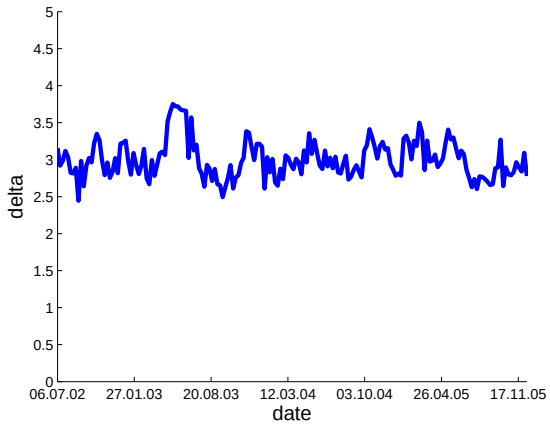


Figure 5.17: Illustration of the size of the uncertainty set over time.

Summarizing the numerical results based on a historical data sample, it can be said that robustification can really lead to an added value in asset management, as it creates more stable portfolios that seem to meet their expectations rather well under normal market conditions.

Chapter 6

Consistency

In Chapters 4 and 5 we have discussed classical and robust portfolio optimization in a practical setting where the necessary parameter estimates are calculated from a finite sample of available historical data. Based on such a finite sample of size S, we have introduced point estimators for the mean vector μ and the covariance matrix Σ of an elliptical distribution. We have furthermore already shown that all the parameter estimators for μ are unbiased. But unbiasedness of the parameter estimates does not yield unbiasedness of the portfolio estimates as the mapping from the set of parameters to optimal portfolios is highly nonlinear.

The intuitive expectation is that the point estimators become more reliable when more data are available for the calculation, i.e. when the sample size S increases. Analogously, we suspect that a larger reliability of the point estimates is reflected in smaller uncertainty sets. Finally, since the optimal solutions of the traditional and the robust portfolio optimization problems based on parameter estimators can be interpreted as estimators for the true portfolio which would be obtained when solving the problem with the original (unknown) market parameters μ and Σ, we furthermore expect that these portfolio estimates are closer to the true portfolio for a larger sample.

In this chapter we thus investigate the behavior of the different estimates (point estimates, uncertainty sets and portfolio estimates) in the case of the sample size S tending to infinity. We will show that all the parameter estimates have the nice property of being *consistent*, expressing that more data lead to more reliable estimators, i.e. for $S \to \infty$, the estimators tend to the true parameter. In mathematical terms, consistency is defined as follows:

Definition 6.1. *Let $\mathcal{Q}_{p,S}$ denote a point estimator for the parameter p based on a sample of size S. The estimator $\mathcal{Q}_{p,S}$ is called*

- weakly consistent *or simply* consistent*, if*

$$\lim_{S\to\infty} \mathbf{P}(|\mathcal{Q}_{p,S} - p| > \varepsilon) = 0,$$

  i.e. if $\mathcal{Q}_{p,S}$ converges in probability to p, denoted by $\mathcal{Q}_{p,S} \xrightarrow{\mathbf{P}} p$,

- strongly consistent, *if*

$$\mathbf{P}(\lim_{S\to\infty} \mathcal{Q}_{p,S} = p) = 1,$$

i.e. if $\mathcal{Q}_{p,S}$ converges almost surely to p, denoted by $\mathcal{Q}_{p,S} \overset{a.s.}{\to} p$.

After summarizing consistency results of the parameter estimates presented in Section 4.3 we extend the concept to uncertainty sets and finally portfolios and investigate these with respect to consistency in Section 6.3. We will show that both the traditional and the robust portfolio estimates are consistent estimates for the true portfolio. The proofs for consistency of the portfolio estimates rely on the fact that the optimal solution of the respective portfolio optimization problem is unique (see Propositions 4.31 and 5.2). This then guarantees continuity of the solution with respect to the parameters according to Theorem 2.45, and a continuous function finally inherits consistency from its arguments.

Throughout this chapter we will assume the same framework as in the previous chapter, i.e. we assume a market of n risky assets whose return vector $R \in \mathbb{R}^n$ follows an elliptical distribution of the form $R \sim \mathcal{E}(\mu, \Sigma, \xi)$. Furthermore, a sample of size S of historical return realizations is supposed to be given and the estimators described in Section 4.3 are considered.

6.1 Consistency of parameter estimates

We first need to establish consistency of the parameter estimators presented in Section 4.3. A property closely related to consistency is *asymptotic normality* which is defined as follows.

Definition 6.2. *Let $\mathcal{Q}_{p,S}$ denote a point estimator for the parameter p based on a sample of size S. The estimator $\mathcal{Q}_{p,S}$ is said to be asymptotically normally distributed with asymptotic covariance matrix K (independent of S), if there is a random variable $Z \sim \mathcal{N}(0, K)$ such that*

$$\sqrt{S}\big(\mathcal{Q}_{p,S} - p\big) \overset{d}{\to} Z \quad \text{for } S \to \infty$$

with "$\overset{d}{\to}$" denoting convergence in distribution.

Before proving asymptotic normality for some of the estimators, we want to state an important property shared by all the L-estimates from above: they depend continuously on the sample $R_1, \ldots, R_S$.

Proposition 6.3. *In the considered framework, all of the L-estimates defined in Section 4.3 are continuous mappings from $(\mathbb{R}^n)^S \to \mathbb{R}^n$.*

Proof. For each component $i = 1, \ldots, n$, the mapping from the original sample $(R_{1,i}, \ldots, R_{S,i}) =: (y_1, \ldots, y_S)$ to the ordered sample $(y_{(1)}, \ldots, y_{(S)})$ is continuous

as it is continuous in each entry. Further, each quantile is continuous in the ordered sample as it is merely a projection onto one element, and finally, each L-estimate is continuous as a linear combination of individual quantiles. □

In the following proposition we show asymptotic normality which – together with unbiasedness – implies the desired property of being consistent estimators for the mean vector μ.

Proposition 6.4. *In the considered framework, all the estimators $\hat{\mu}_S^{ML}$, $\hat{\mu}_S^{ME}$, $\hat{\mu}_S^{QR}$, $\hat{\mu}_S^{TM}$ and $\hat{\mu}_S^{HU}$ for the mean vector μ are asymptotically normally distributed.*

Proof.

- Due to the existence of continuous marginal densities, each finite collection of quantiles asymptotically follows a multivariate normal distribution, see Shorack and Wellner [78], Example 1 on page 639. Thus, both the median and the quartile estimator are asymptotically normally distributed.
- The trimmed mean is asymptotically normal by Theorem 3.2 on page 60 in Huber [38], as all marginal distributions have continuous densities.
- According to Huber [37], Section 4, the maximum likelihood estimator is asymptotically normal.
- Following Huber [38], Theorem 2.4, page 50, asymptotic normality is given for the Huber estimator as the marginal densities are sufficiently smooth in our framework. □

Based on asymptotic normality, it is easy to derive consistency of the estimators.

Theorem 6.5. *In the given setting it holds that $\hat{\mu}_S^{ML}$, $\hat{\mu}_S^{ME}$, $\hat{\mu}_S^{QR}$, $\hat{\mu}_S^{TM}$ and $\hat{\mu}_S^{HU}$ are consistent estimators for μ. Further, $\hat{\Sigma}_S^{ML}$ is a consistent estimator for Σ.*

Proof. From asymptotic normality, convergence in distribution to μ is straightforward as the estimators are unbiased (the covariance matrix of Z divided by $\sqrt{S}$ tends to zero for $S \to \infty$). Furthermore, since the limit μ is a constant, we also obtain convergence in probability, i.e. consistency, see e.g. Jacod and Protter [41], Theorem 18.3. Thus, it only remains to show consistency of the maximum likelihood estimator for Σ. This follows immediately by Rinne [69], page 454. □

For our selected estimators, it is even possible to show strong consistency. The following proposition summarizes those results.

Proposition 6.6. *In the given setting, the estimators $\hat{\mu}_S^{ML}$, $\hat{\mu}_S^{ME}$, $\hat{\mu}_S^{QR}$, $\hat{\mu}_S^{TM}$ and $\hat{\mu}_S^{HU}$ are strongly consistent estimators for μ. Further, $\hat{\Sigma}_S^{ML}$ is a strongly consistent estimator for Σ.*

Proof. Corollary 2.2 of Chapter 3 in Huber [38] gives strong consistency of M-estimates, if the limit point is unique – which is the case for elliptical distributions. Hence, the mean, the median and the Huber estimator are strongly consistent estimators for μ in our setting. Furthermore, $\hat{\Sigma}_S^{ML}$ is a maximum likelihood type estimator for Σ and thus falls into that category as well.

For the L-estimators $\hat{\mu}_S^{QR}$ and $\hat{\mu}_S^{TM}$ we obtain strong consistency by Theorem 3.1 of Chapter 3 and Proposition 6.1 of Chapter 2 in Huber [38]. □

Table 6.1 collects the properties of the presented parameter estimates.

	unbiased	asymp. unbiased	asympt. normal	consistent	strongly consistent
Mean	yes	yes	yes	yes	yes
Median	yes	yes	yes	yes	yes
Quartile	yes	yes	yes	yes	yes
Trimmed	yes	yes	yes	yes	yes
Huber	yes	yes	yes	yes	yes
MLE for Σ	no	yes	—	yes	yes

Table 6.1: Properties of the selected estimators.

6.2 Consistency of uncertainty sets

In Section 6.3 we want to apply the concept of consistency to portfolio estimators, both in the classical and the robust portfolio optimization setting. As in the robust problem not only point estimates are needed, but entire sets containing possible parameter realizations come into play, we need to introduce the notion of consistency for uncertainty sets by considering them to be set-valued random variables.

Definition 6.7. *An uncertainty set $\mathcal{U}_S$ is called*

- weakly consistent *or simply* consistent *for the pair of parameters (μ, Σ) if*

$$H_d\big(\mathcal{U}_S, \{(\mu, \Sigma)\}\big) \to 0 \quad \textit{in probability for } S \to \infty,$$

- strongly consistent *for the pair of parameters (μ, Σ) if*

$$H_d\big(\mathcal{U}_S, \{(\mu, \Sigma)\}\big) \to 0 \quad \textit{almost surely for } S \to \infty,$$

with $H_d(A, B)$ denoting the Hausdorff distance between the sets A and B as defined in Appendix B.

The property of consistency of an uncertainty set naturally depends on the particular definition in the practical application. In Chapter 5 we have presented two uncertainty sets: the first one was defined by a confidence ellipsoid, i.e. the distribution (the first two moments, respectively) of the uncertain parameter was needed; the second one was built as the smallest ellipsoid containing a number of (statistical) point estimates for the parameter of interest. Before showing in the subsequent propositions that these two types of uncertainty sets are consistent, we shortly recall their definitions, written here in the more general form as a joint uncertainty set for (μ, Σ) even though we only explicitly account for uncertainty of μ. The dependence on S of the uncertainty set and the parameter estimates therein is pointed out by the subscript S.

- The distribution of μ is supposed to be given by $\mu \sim \mathcal{E}(\hat{\mu}_S, \frac{1}{S}\hat{\Sigma}_S, \phi)$ with $\hat{\mu}_S$ and $\hat{\Sigma}_S$ being (strongly) consistent estimators for μ and Σ, respectively. The uncertainty set defined as a confidence ellipsoid is thus given by

$$\begin{aligned}\mathcal{U}_{S,\text{conf}}(\hat{\mu}_S, \hat{\Sigma}_S) &= \{(r, C) \mid (r - \hat{\mu}_S)^T \big(\frac{1}{S}\hat{\Sigma}_S\big)^{-1}(r - \hat{\mu}_S) \leq \delta^2, C = \hat{\Sigma}_S\} \\ &= \{r \mid (r - \hat{\mu}_S)^T \hat{\Sigma}_S^{-1}(r - \hat{\mu}_S) \leq \frac{\delta^2}{S}\} \times \{\hat{\Sigma}_S\}. \end{aligned} \tag{6.1}$$

- A different approach to defining an uncertainty set is by considering a set M of finitely many (strongly) consistent estimators for the parameter μ. Furthermore, let $\hat{\Sigma}_S$ be a (strongly) consistent estimator for Σ. The uncertainty set for μ is then described by the smallest ellipsoid containing all points of M, thus we obtain

$$\mathcal{U}_{S,M} = \{(r, C) \mid (r - \bar{\mu}_S)^T \bar{\Sigma}_S^{-1}(\mu - \bar{\mu}_S) \leq \delta_S^2, C = \hat{\Sigma}_S\} \tag{6.2}$$

  with

$$\begin{aligned}\bar{\mu}_S &= \frac{1}{|M|}\sum_{m \in M} m \\ \bar{\Sigma}_S &= \text{diag}(\bar{\sigma}_{S,11}^2, \ldots, \bar{\sigma}_{S,nn}^2) \quad \text{where} \quad \bar{\sigma}_{S,ii}^2 = \frac{1}{|M| - 1}\sum_{m \in M}(m_i - \bar{\mu}_{S,i})^2 \\ \delta_S^2 &= \max_{m \in M}(m - \bar{\mu}_S)^T \bar{\Sigma}_S^{-1}(m - \bar{\mu}_S).\end{aligned}$$

In the following propositions we will show consistency of the two general types of uncertainty sets just described. We omit the explicit analysis of the joint confidence ellipsoid for μ and Σ since the focus is on uncertainty sets for the return vector only.

Proposition 6.8. *Let $\hat{\mu}_S$ and $\hat{\Sigma}_S$ be (strongly) consistent estimators for μ and Σ, respectively. Then, the uncertainty set $\mathcal{U}_{S,conf}(\hat{\mu}_S, \hat{\Sigma}_S)$ as given in Equation (6.1) is (strongly) consistent for (μ, Σ).*

Proof. To prove (strong) consistency, we need to show that the Hausdorff distance between the uncertainty set $\mathcal{U}_S := \mathcal{U}_{S,\text{conf}}(\hat{\mu}_S, \hat{\Sigma}_S)$ and the singleton set $\{(\mu, \Sigma)\}$ tends to zero. Using the definition of the Hausdorff distance (see Definition B.1) and noting that the excess of $\mathcal{U}_S$ over $\{(\mu, \Sigma)\}$ is always greater than the excess of $\{(\mu, \Sigma)\}$ over $\mathcal{U}_S$, we obtain

$$\begin{aligned}
H_d\big(\mathcal{U}_S, \{(\mu,\Sigma)\}\big) &= \max\{e_d\big(\mathcal{U}_S, \{(\mu,\Sigma)\}\big), e_d\big(\{(\mu,\Sigma)\}, \mathcal{U}_S\big)\} \\
&= e_d\big(\mathcal{U}_S, \{(\mu,\Sigma)\}\big) \\
&= \sup_{(r,C)\in\mathcal{U}_S} d\big((r,C),(\mu,\Sigma)\big) \\
&\overset{(C.3)}{=} \sup_{(r,C)\in\mathcal{U}_S} \|r-\mu\|_2 + \|C-\Sigma\|_{tr} \\
&\overset{C=\hat{\Sigma}_S}{=} \sup_{(r,C)\in\mathcal{U}_S} \|r-\mu\|_2 + \|\hat{\Sigma}_S-\Sigma\|_{tr} \\
&\leq \sup_{(r,C)\in\mathcal{U}_S} \|r-\hat{\mu}_S\|_2 + \underbrace{\|\hat{\mu}_S-\mu\|_2}_{\to 0} + \underbrace{\|\hat{\Sigma}_S-\Sigma\|_{tr}}_{\to 0}.
\end{aligned}$$

The expressions $\|\hat{\mu}_S-\mu\|_2$ and $\|\hat{\Sigma}_S-\Sigma\|_{tr}$ tend to zero due to (strong) consistency of the respective estimators. It remains to show that $\|r-\hat{\mu}_S\|_2 \to 0$, i.e. $r \to \hat{\mu}_S$ for $S \to \infty$. Recall that an equivalent formulation of $\mathcal{U}_S$ is given by (see Example 3.26)

$$\mathcal{U}_S = \{r \mid r = \hat{\mu}_S + \frac{\delta}{\sqrt{S}}\hat{\Sigma}_S^{\frac{1}{2}} z, \|z\|_2 \leq 1\} \times \{\hat{\Sigma}_S\}.$$

To prove that $\mathcal{U}_S$ eventually reduces to a single point, it suffices to show that $\frac{\delta}{\sqrt{S}} \cdot \hat{\Sigma}_S^{\frac{1}{2}} \to 0$ which holds if $\hat{\Sigma}_S^{\frac{1}{2}}$ or equivalently $\hat{\Sigma}_S$ is bounded, since $\frac{\delta}{\sqrt{S}} \to 0$ for $S \to \infty$. As the maximum likelihood estimator $\hat{\Sigma}_S$ is given by (see Definition 4.15)

$$\hat{\Sigma}_S = \frac{1}{S}\sum_{s=1}^{S}(R_s-\hat{\mu}_S)(R_s-\hat{\mu}_S)^T,$$

it is naturally bounded for sufficiently large S. Thus, it holds that $r \to \hat{\mu}_S$ and finally

$$H_d\big(\mathcal{U}_{S,M}, \{(\mu,\Sigma)\}\big) \to 0,$$

i.e. (strong) consistency of $\mathcal{U}_S$ for (μ, Σ) is proved. □

Proposition 6.9. *Let M be a set of finitely many (strongly) consistent estimators for μ, and let $\hat{\Sigma}_S$ be a (strongly) consistent estimator for Σ. Then, the uncertainty set $\mathcal{U}_{S,M}$ as given in Equation (6.2) is (strongly) consistent for (μ, Σ).*

Proof. First note that since M contains only (strongly) consistent estimators for μ, this also implies that $\bar{\mu}_S$ as the mean over the elements in M is a (strongly) consistent estimator for μ. We now proceed as in the proof of Proposition 6.8:

$$\begin{aligned} H_d\big(\mathcal{U}_{S,M}, \{(\mu,\Sigma)\}\big) &= e_d\big(\mathcal{U}_{S,M}, \{(\mu,\Sigma)\}\big) \\ &= \sup_{(r,C)\in\mathcal{U}_{S,M}} d\big((r,C),(\mu,\Sigma)\big) \\ &\overset{(C.3)}{=} \sup_{(r,C)\in\mathcal{U}_{S,M}} \|r-\mu\|_2 + \|\hat{\Sigma}_S - \Sigma\|_{tr} \\ &\leq \sup_{(r,C)\in\mathcal{U}_{S,M}} \|r-\bar{\mu}_S\|_2 + \underbrace{\|\bar{\mu}_S-\mu\|_2}_{\to 0} + \underbrace{\|\hat{\Sigma}_S-\Sigma\|_{tr}}_{\to 0} \end{aligned}$$

where convergence is meant either in probability or almost surely, depending on the prerequisites. It thus remains to show that $r \to \bar{\mu}_S$ for $S \to \infty$. We again use the equivalent formulation of $\mathcal{U}_{S,M}$ which is given by

$$\mathcal{U}_{S,M} = \{r \mid r = \bar{\mu}_S + \delta_S \bar{\Sigma}_S^{\frac{1}{2}} z, \|z\|_2 \leq 1\} \times \{\hat{\Sigma}_S\}.$$

To prove that $\mathcal{U}_{S,M}$ eventually reduces to a single point, it suffices to show that $\bar{\Sigma}_S$ tends to the zero matrix and δ_S is bounded, thus $\delta_S \cdot \bar{\Sigma}_S \to 0$ and hence, $r \to \bar{\mu}_S$. Convergence of the diagonal entries $\bar{\sigma}_{S,ii}$ to zero follows directly from consistency of the individual estimators since convergence (almost surely or in probability) of parameters transfers to continuous functions thereof, see e.g. Jacod and Protter [41], Theorem 17.5. For the sizing variable δ_S it holds that

$$\begin{aligned} \delta_S^2 &= \max_{m\in M} (m-\bar{\mu}_S)^T \bar{\Sigma}_S^{-1} (m-\bar{\mu}_S) \\ &= \max_{m\in M} \sum_{i=1}^{n} (m_i - \bar{\mu}_{S,i})^2 \cdot \frac{1}{\bar{\sigma}_{S,ii}^2} \\ &= \max_{m\in M} \sum_{i=1}^{n} \frac{(m_i-\bar{\mu}_{S,i})^2}{\frac{1}{|M|-1}\sum_{l\in M}(l_i-\bar{\mu}_{S,i})^2} \\ &= \max_{m\in M} (|M|-1) \sum_{i=1}^{n} \underbrace{\frac{(m_i-\bar{\mu}_{S,i})^2}{(m_i-\bar{\mu}_{S,i})^2 + \sum_{l\in M, l\neq m}(l_i-\bar{\mu}_{S,i})^2}}_{\leq 1} \\ &\leq (|M|-1)\, n < \infty. \end{aligned}$$

Thus, altogether, we have $\delta_S \cdot \bar{\Sigma}_S \to 0$ and hence $r \to \bar{\mu}_S$ which finally yields

$$H_d\big(\mathcal{U}_{S,M}, \{(\mu,\Sigma)\}\big) \to 0,$$

i.e. (strong) consistency of $\mathcal{U}_{S,M}$ for (μ, Σ). □

After having established consistency of the uncertainty sets under consideration, we extend the concept of consistency to portfolio estimates in the following section.

6.3 Consistency of portfolio estimates

In this section we investigate classical and robust portfolios, i.e. the solutions of the respective optimization problems, with respect to the property *consistency*. This is done by interpreting an optimal portfolio obtained as solution of the portfolio optimization problem based on parameter estimates $\hat{\mu}$ and $\hat{\Sigma}$ as estimates for the true optimal portfolio, the solution of the problem with the original parameters μ and Σ.

To set and clarify the notation, we recall the traditional and the robust portfolio optimization problem. For exclusion of any degenerate cases and to assure existence and uniqueness of the optimal solution, we make the following two assumptions which are not very strong restrictions in practice.

Assumption 6.10. *Let the set of feasible portfolios be described by a non-empty, convex and compact set X such that*

$$X \subset \{x \in \mathbb{R}^n \mid x^T \mathbf{1} = 1\}.$$

Assumption 6.11. *Throughout the section we assume that both the classical and the robust portfolio optimization problem possesses a unique solution for $\lambda = 1$.*

Recall that for both the classical and the robust portfolio optimization problem, we have shown uniqueness of the optimal solution for $\lambda < 1$ in Propositions 4.31 and 5.2.

The classical portfolio optimization problem based on the true – but unknown – parameters μ and Σ is described by

$$\min_{x \in X} \quad (1-\lambda)\sqrt{x^T \Sigma x} - \lambda x^T \mu. \tag{$P_{\lambda,\mu,\Sigma}$}$$

The optimal solution for a particular trade-off level $\lambda \in [0,1]$ will be denoted by $x^*(\lambda) := x^*(\lambda, \mu, \Sigma) = x^*_{cl}(\lambda, \mu, \Sigma)$. Solving this optimization problem using estimators $\hat{\mu}_S$ and $\hat{\Sigma}_S$ instead of μ and Σ, the solution will be denoted by $x^*_{cl}(\lambda, \hat{\mu}_S, \hat{\Sigma}_S)$ where the subscript S at the parameters expresses their dependence on the sample size. With the subscript "cl" we distinguish more clearly this classical or traditional portfolio from the robust one below.

Next we recall the robust optimization problem, described in terms of an uncertainty set $\mathcal{U}$.

$$\min_{x \in X} \max_{(r,C) \in \mathcal{U}} \quad (1-\lambda)\sqrt{x^T C x} - \lambda x^T r. \tag{$RP_{\lambda,\mathcal{U}}$}$$

Also in this case we consider the problem $(RP_{\lambda,\mathcal{U}_S})$ where parameter estimates are used to determine the uncertainty set $\mathcal{U}_S$. Again, the subscript S in the notation $\mathcal{U}_S$ represents the (indirect) dependence of the uncertainty set on the sample of size S, from which the parameters describing $\mathcal{U}_S$ are calculated. The optimal solution of the robust problem will accordingly be denoted by $x^*_{rob}(\lambda,\mathcal{U}_S)$.

Both of these portfolio estimates $x^*_{cl}(\lambda,\hat{\mu}_S,\hat{\Sigma}_S)$ and $x^*_{rob}(\lambda,\mathcal{U}_S)$ can be interpreted as estimators for the true optimal portfolio $x^*(\lambda) = x^*(\lambda,\mu,\Sigma)$ and we want to show consistency thereof in the following.

Theorem 6.12. *Let $0 \leq \lambda \leq 1$. Then the classical optimal portfolio $x^*_{cl}(\lambda,\hat{\mu}_S,\hat{\Sigma}_S)$ is a (strongly) consistent estimator for the true portfolio $x^*(\lambda)$ if the parameter estimators $\hat{\mu}_S$ and $\hat{\Sigma}_S$ are (strongly) consistent estimators for μ and Σ.*

Proof. Due to uniqueness of the optimal solution, continuity of $x^*_{cl}(\lambda,\hat{\mu}_S,\hat{\Sigma}_S)$ for all parameters $(\hat{\mu}_S,\hat{\Sigma}_S)$ is given by Theorem 2.45 as X is constant and thus Hausdorff continuous. Furthermore, according to Jacod and Protter [41], Theorem 17.5, the proof concludes as convergence (in probability or almost surely) of random variables transfers to continuous functions thereof. □

Next, we show that the robust portfolio estimator calculated using an uncertainty set which depends on the sample size S tends (in probability or almost surely) to the portfolio obtained when solving the classical problem using the true market parameters. This is intuitively expected since when considering an uncertainty set described by parameters that become more and more exact, the uncertainty set gets smaller and smaller and finally shrinks to a point only – the true parameters.

Theorem 6.13. *Let $0 \leq \lambda \leq 1$ and let $\mathcal{U}_S$ be a (strongly) consistent uncertainty set for (μ,Σ). Then the robust portfolio estimator $x^*_{rob}(\lambda,\mathcal{U}_S)$ is a (strongly) consistent estimator for the true optimal portfolio $x^*(\lambda)$.*

Proof. The proof will follow analogous lines as the proof for the consistency of the classical portfolio, see Theorem 6.12. Hence, we need continuity of the optimal solution with respect to $\mathcal{U}_S$, then the proof concludes again by the transfer of consistency of $\mathcal{U}_S$ to a continuous function thereof, see Jacod and Protter [41], Theorem 17.5.

As we have assumed uniqueness of the optimal solution $x^*_{rob}(\lambda,\mathcal{U})$, it suffices to show continuity of the robust objective function in $\mathcal{U}$ ($\mathcal{U}$ arbitrary) to obtain continuity of $x^*_{rob}(\lambda,\mathcal{U})$ according to Theorem 2.45.

Thus, we first show that the mapping $f_{rob} : \mathbb{R}^n \times [0,1] \times \mathcal{U} \to \mathbb{R}$ with

$$f_{rob}(x,\lambda,\mathcal{U}) := \max_{(r,C)\in\mathcal{U}} \quad (1-\lambda)\sqrt{x^TCx} - \lambda x^Tr$$

is continuous in $\mathcal{U}$ for all $x \in X$ and $\lambda \in [0,1]$. For notational convenience we introduce the classical objective function as

$$f_{cl}(x,\lambda,r,C) := (1-\lambda)\sqrt{x^TCx} - \lambda(x^Tr).$$

Thus, the robust objective f_{rob} can also be expressed as

$$f_{rob}(x, \lambda, \mathcal{U}) := \max_{(r,C)\in\mathcal{U}} \quad f_{cl}(x, \lambda, r, C).$$

Now, let a sequence of uncertainty sets $\{\mathcal{U}_n\}$ with $\mathcal{U}_n \to \mathcal{U}$ be given, where the convergence is to be understood in the Hausdorff sense, i.e. $\lim_{n\to\infty} H_d(\mathcal{U}_n, \mathcal{U}) = 0$. We show

$$\lim_{n\to\infty} f_{rob}(x, \lambda, \mathcal{U}_n) = f_{rob}(x, \lambda, \mathcal{U})$$

by showing both inequalities.

(i) Let (r^*, C^*) be a maximizing element in $\mathcal{U}$ i.e. (r^*, C^*) maximizes f_{cl} for x and λ, formally written as

$$f_{cl}(x, \lambda, r^*, C^*) = \max_{(r,C)\in\mathcal{U}} f_{cl}(x, \lambda, r, C) = f_{rob}(x, \lambda, \mathcal{U}).$$

Due to convergence of $\mathcal{U}_n$ to $\mathcal{U}$ there exist pairs $(r_n, C_n) \in \mathcal{U}_n$ with $(r_n, C_n) \to (r^*, C^*)$. Furthermore, using continuity of $f_{cl}(x, \lambda, r, C)$ with respect to (r, C), we get

$$\begin{aligned} f_{rob}(x, \lambda, \mathcal{U}) &= f_{cl}(x, \lambda, r^*, C^*) \\ &= \lim_{n\to\infty} f_{cl}(x, \lambda, r_n, C_n) \\ &\leq \lim_{n\to\infty} \max_{(r,C)\in\mathcal{U}_n} f_{cl}(x, \lambda, r, C) \\ &= \lim_{n\to\infty} f_{rob}(x, \lambda, \mathcal{U}_n). \end{aligned}$$

(ii) Let $(r_n^*, C_n^*) \in \mathcal{U}_n$ be the maximizing elements for $\mathcal{U}_n$, i.e.

$$f_{cl}(x, \lambda, r_n^*, C_n^*) = \max_{(r,C)\in\mathcal{U}_n} f_{cl}(x, \lambda, r, C) = f_{rob}(x, \lambda, \mathcal{U}_n).$$

Due to Hausdorff convergence of $\mathcal{U}_n$ to $\mathcal{U}$ there exist $(r_n, C_n) \in \mathcal{U}$ with $||r_n - r_n^*||_2 \to 0$ and $||C_n - C_n^*||_{tr} \to 0$. Then any accumulation point $(r^*, C^*) \in \mathcal{U}$ is also an accumulation point for the sequence $\{(r_n^*, C_n^*)\}$. Without loss of generality we assume that this accumulation point is unique. Then we get

$$\begin{aligned} \lim_{n\to\infty} f_{rob}(x, \lambda, \mathcal{U}_n) &= \lim_{n\to\infty} f_{cl}(x, \lambda, r_n^*, C_n^*) \\ &= f_{cl}(x, \lambda, r^*, C^*) \\ &\leq \max_{(r,C)\in\mathcal{U}} f_{cl}(x, \lambda, r, C) \\ &= f_{rob}(x, \lambda, \mathcal{U}). \end{aligned}$$

Hence, we now have established

$$\lim_{n\to\infty} f_{rob}(x,\lambda,\mathcal{U}_n) = f_{rob}(x,\lambda,\mathcal{U}),$$

i.e. continuity of f_{rob} with respect to $\mathcal{U}$.

Continuity of the objective function f_{rob} (in x and $\mathcal{U}$) together with a constant feasibility set X and a unique solution yields continuity $x^*_{rob}(\lambda,\mathcal{U}_S)$ in $\mathcal{U}_S$ by Theorem 2.45. □

Chapter 7

Portfolio optimization under uncertainty and prior knowledge

As discussed in Chapters 4 and 5 we need *point estimates* for the unknown market parameters μ and Σ to be able to solve the classical portfolio optimization or an *uncertainty set* for the robust optimization. In most applications, simply the maximum likelihood estimators based on a historical data sample are used to replace the true parameters in the classical optimization, but there are a number of different statistical estimators which could be used instead, see Section 4.3.

In financial practice it is often wanted to take individual opinions about the future performance of some assets into account, e.g. a stock index manager could have a rather precise idea about the development of that particular index – which could be different from the expectation obtained by using the historical data as a reference. Hence, concepts combining both external knowledge and estimates based on a data sample like the Bayesian model and the Black-Litterman approach (Black and Litterman [15]) gained more and more interest in recent years.

In this chapter we recall in great detail the two approaches to obtain point estimates for the classical portfolio optimization problem. Furthermore, their applicability for determining uncertainty sets for the robust portfolio optimization problem is studied. Besides considering the Bayes model with a continuous prior which was already done by Meucci in [57], we also present possibilities to define an uncertainty set in the Bayes model with a discrete prior and in the Black-Litterman framework.

Furthermore, in Sections 7.3 and 7.4 we analyze if or under which conditions one of the models can be seen as a special case of the other and we compare the estimates and uncertainty sets obtained thereof.

In the Bayesian approach (see e.g. Meucci [57]) in Section 7.1 a prior distributional assumption about the parameters to be estimated is made, e.g. by some expert. Then, this prior is conditioned on the available data sample to obtain the final estimates.

The Black-Litterman approach (Black and Litterman [15]) illustrated in Sec-

tion 7.2 works somehow differently as it uses the data sample to describe the prior distribution and then incorporates explicit investor forecasts on individual asset performances to determine the combined estimate.

7.1 Bayesian approach

Before applying the Bayesian approach in the particular portfolio optimization framework where estimates for the two unknown parameters μ and Σ need to be determined, we give a description of the general methodology. Subsuming all considered parameters into a general one, denoted by θ (i.e. in our case, θ represents μ and Σ), the procedure to determine Bayesian estimates and / or an uncertainty set is given as follows:

1. We make some *prior* assumption about the distribution of the unknown parameters, i.e. we assume θ to have a density function described by $\varphi_{\text{prior}}(\theta)$.

2. We obtain additional market information in form of a sample X=$X_1, ...,X_S$ which was drawn according to a given distribution depending on the (yet unknown) parameter θ. The joint density function of the sample X is thus denoted by $\varphi(X \mid \theta)$.

3. The *posterior* distribution finally gives the distribution of the parameters after consideration of the additional market information – which is what we are eventually looking for. According to the Bayes rule, the posterior density calculates to

$$\begin{aligned}\varphi_{\text{post}}(\theta \mid X) &= \frac{\varphi(X,\theta)}{\varphi(X)} \\ &= \frac{\varphi(X \mid \theta)\varphi_{\text{prior}}(\theta)}{\int \varphi(X \mid \beta)\varphi_{\text{prior}}(\beta)d\beta} \\ &= \gamma\varphi(X \mid \theta)\varphi_{\text{prior}}(\theta)\end{aligned}$$

   with γ a suitable normalizing constant, i.e.

$$\gamma := \left(\int \varphi(X \mid \theta)\varphi_{\text{prior}}(\theta)d\theta\right)^{-1}.$$

Generally, a point estimate $\hat{\theta}$ for θ can be obtained by minimizing the expected loss, i.e. by solving

$$\min_{\hat{\theta}} \quad \mathbf{E}[\|\theta - \hat{\theta}\|_2^2].$$

Equivalently expressed, we solve

$$\begin{aligned} &\min_{\hat{\theta}} \quad \mathbf{E}[\|\theta - \hat{\theta}\|_2^2] \\ =&\min_{\hat{\theta}} \quad \mathbf{E}[\theta^T\theta - 2\theta^T\hat{\theta} + \hat{\theta}^T\hat{\theta}] \\ =&\min_{\hat{\theta}} \quad \mathbf{E}[\theta^T\theta] - 2\,\mathbf{E}[\theta]^T\hat{\theta} + \hat{\theta}^T\hat{\theta} \end{aligned}$$

which is a quadratic function in $\hat{\theta}$. From setting the derivative equal to zero, we get

$$\begin{aligned} -2\,\mathbf{E}[\theta] + 2\hat{\theta} &\stackrel{!}{=} 0 \\ \Rightarrow \quad \hat{\theta} &= \mathbf{E}[\theta]. \end{aligned} \tag{7.1}$$

From the posterior distribution given above, we can obtain both the Bayesian point estimate and an uncertainty set. The point estimate in the Bayesian approach is thus given by

$$\hat{\theta} := \mathbf{E}[\theta \mid X]$$

and an uncertainty set can be created by using the first two moments of the posterior distribution of $\theta \mid X$ to form a confidence ellipsoid.

Remark 7.1. *Instead of using the point estimate coming from minimizing the expected loss, we could also use the maximum likelihood estimator obtained from the posterior distribution.*

7.1.1 Bayesian approach with a continuous prior

After having a general description of the Bayesian approach, we now perform the individual steps in more detail using a particular prior. These results can also be found in Meucci [57], but we will nevertheless state the explicit calculations for completeness. We assume that (μ, Σ) follows a normal inverse Wishart ($\mathcal{NIW}$) distribution, see Appendix D.3. This is the most natural distribution if we want to assume variability both for the vector of expected returns and for the covariance matrix. Often, a distributional assumption (mostly the normal distribution) is only made for the return vector while the covariance matrix is assumed to be fix. We will encounter such a framework below when discussing the Black-Litterman approach.

In the following calculations to obtain the Bayesian point estimates, $\gamma_j, j \in \mathbb{N}$ are supposed to denote normalizing constants subsuming all leftover non-relevant expressions (like e.g. $(2\pi)^{-\frac{n}{2}}$) and chosen appropriately such that the respective function represents a probability density.

1. *Prior assumption*
   The prior assumption about the distribution of the parameters is the following:

$$(\mu, \Sigma) \sim \mathcal{NIW}(\mu_0, d_0, \Sigma_0, \nu_0)$$

   with the joint density (see Formula D.2)

$$\begin{aligned}\varphi_{\text{prior}}(\mu, \Sigma) &= \varphi_{\mathcal{NIW}}(\mu, \Sigma) \\ &= \gamma_1 |\Sigma^{-1}|^{\frac{\nu_0+n+2}{2}} \exp\left\{-\frac{1}{2}[d_0(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0) + \operatorname{tr}(\nu_0\Sigma_0 \cdot \Sigma^{-1})]\right\}.\end{aligned}$$

2. *Market information*
   The market information is collected within the sample realizations $x_1, \ldots, x_S$ with $X_i \mid \mu, \Sigma \sim \mathcal{N}(\mu, \Sigma), \quad i = 1, \ldots, S$ i.i.d. Thus, we can calculate the joint probability density function φ_M of the entire sample as follows:

$$\begin{aligned}\varphi_M(x_1, \ldots, x_S \mid \mu, \Sigma) &= \prod_{s=1}^{S} \varphi(x_s \mid \mu, \Sigma) \\ &= \prod_{s=1}^{S} \frac{1}{(2\pi)^{\frac{n}{2}}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x_s-\mu)^T\Sigma^{-1}(x_s-\mu)\right\} \\ &= \left(\frac{1}{(2\pi)^{\frac{n}{2}}}\right)^S |\Sigma|^{-\frac{S}{2}} \exp\left\{-\frac{1}{2}\sum_{s=1}^{S}(x_s-\mu)^T\Sigma^{-1}(x_s-\mu)\right\}.\end{aligned}$$

   With $\hat{\mu}$ and $\hat{\Sigma}$ denoting the maximum likelihood estimators of the sample of realizations $x_1, \ldots, x_S$, i.e.

$$\begin{aligned}\hat{\mu} &= \frac{1}{S}\sum_{s=1}^{T} x_s, \\ \hat{\Sigma} &= \frac{1}{S}\sum_{s=1}^{T}(x_s-\hat{\mu})(x_s-\hat{\mu})^T,\end{aligned}$$

   we can refomulate the sum in the exponential function as

$$\begin{aligned}&\sum_{s=1}^{S}(x_s-\mu)^T\Sigma^{-1}(x_s-\mu) = \\ &\quad= \sum_{s=1}^{S}\left(x_s^T\Sigma^{-1}x_s - 2x_s\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu\right) \\ &\quad= S\left[\mu^T\Sigma^{-1}\mu - 2\hat{\mu}^T\Sigma^{-1}\mu\right] + \sum_{s=1}^{S} x_s^T\Sigma^{-1}x_s\end{aligned}$$

$$
\begin{aligned}
&= S\big[(\hat{\mu}-\mu)^T\Sigma^{-1}(\hat{\mu}-\mu)\big] - \underbrace{S\,\hat{\mu}^T\Sigma^{-1}\hat{\mu}}_{=\mathrm{tr}(S\hat{\mu}\hat{\mu}^T\Sigma^{-1})} + \underbrace{\sum_{s=1}^{S} x_s^T\Sigma^{-1}x_s}_{=\mathrm{tr}\left(\sum_{s=1}^{S} x_s x_s^T\Sigma^{-1}\right)} \\
&= S(\hat{\mu}-\mu)^T\Sigma^{-1}(\hat{\mu}-\mu) + \mathrm{tr}\left(\left[\sum_{s=1}^{S} x_s x_s^T - S\hat{\mu}\hat{\mu}^T\right]\Sigma^{-1}\right) \\
&= S(\hat{\mu}-\mu)^T\Sigma^{-1}(\hat{\mu}-\mu) + \mathrm{tr}\left(\sum_{s=1}^{S}(x_s-\hat{\mu})(x_s-\hat{\mu})^T\Sigma^{-1}\right) \\
&= S(\hat{\mu}-\mu)^T\Sigma^{-1}(\hat{\mu}-\mu) + \mathrm{tr}\big(S\hat{\Sigma}\Sigma^{-1}\big)
\end{aligned}
$$

and thus, we finally obtain the probability density function of the distribution of the sample:

$$
\begin{aligned}
&\varphi_M(x_1,\ldots,x_S \mid \mu,\Sigma) = \qquad (7.2)\\
&\quad = \left(\frac{1}{(2\pi)^{\frac{n}{2}}}\right)^S |\Sigma|^{-\frac{S}{2}} \exp\left\{-\frac{1}{2}[S(\hat{\mu}-\mu)^T\Sigma^{-1}(\hat{\mu}-\mu) + \mathrm{tr}(S\hat{\Sigma}\Sigma^{-1})]\right\}.
\end{aligned}
$$

3. *Posterior distribution*
We can now calculate the posterior density of the parameters (μ,Σ) conditioned on the additional market information:

$$
\begin{aligned}
&\varphi_{\mathrm{post}}(\mu,\Sigma \mid x_1,\ldots,x_S) = \gamma_2\varphi_M(x_1,\ldots,x_S \mid \mu,\Sigma)\varphi_{\mathrm{prior}}(\mu,\Sigma) \\
&\quad = \gamma_2\left(\frac{1}{(2\pi)^{\frac{n}{2}}}\right)^S |\Sigma|^{-\frac{S}{2}} \exp\left\{-\frac{1}{2}[S(\hat{\mu}-\mu)^T\Sigma^{-1}(\hat{\mu}-\mu) + \mathrm{tr}(S\hat{\Sigma}\Sigma^{-1})]\right\} \\
&\qquad \cdot \gamma_1\,|\Sigma^{-1}|^{\frac{\nu_0+n+2}{2}} \exp\left\{-\frac{1}{2}[d_0(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0) + \mathrm{tr}(\nu_0\Sigma_0\cdot\Sigma^{-1})]\right\} \\
&\quad = \gamma_3\,|\Sigma|^{-\frac{(\nu_0+S)+n+2}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left((S\hat{\Sigma}+\nu_0\Sigma_0)\cdot\Sigma^{-1}\right)\right\} \\
&\qquad \cdot \exp\left\{-\frac{1}{2}[S(\hat{\mu}-\mu)^T\Sigma^{-1}(\hat{\mu}-\mu) + d_0(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)]\right\} \\
&\quad \overset{(*)}{=} \gamma_3\,|\Sigma|^{-\frac{(\nu_0+S)+n+2}{2}} \\
&\qquad \cdot \exp\left\{-\frac{1}{2}\left[(d_0+S)\left(\mu-\frac{S\hat{\mu}+d_0\mu_0}{d_0+S}\right)^T\Sigma^{-1}\left(\mu-\frac{S\hat{\mu}+d_0\mu_0}{d_0+S}\right) + \right.\right. \\
&\qquad \left.\left. +\ \mathrm{tr}\left(\left[S\hat{\Sigma}+\nu_0\Sigma_0+\frac{d_0S}{d_0+S}(\hat{\mu}-\mu_0)(\hat{\mu}-\mu_0)^T\right]\Sigma^{-1}\right)\right]\right\} \\
&\quad = \gamma_3|\Sigma|^{-\frac{\nu_1+n+2}{2}} \exp\left\{-\frac{1}{2}\left[d_1(\mu-\mu_1)^T\Sigma^{-1}(\mu-\mu_1) + \mathrm{tr}(\nu_1\Sigma_1\Sigma^{-1})\right]\right\}
\end{aligned}
$$

with the parameters

$$\begin{aligned}
\nu_1 &= \nu_0 + S,\\
d_1 &= d_0 + S,\\
\mu_1 &= \frac{d_0\mu_0 + S\hat{\mu}}{d_0 + S},\\
\Sigma_1 &= \frac{1}{\nu_0 + S}\left[S\hat{\Sigma} + \nu_0\Sigma_0 + \frac{d_0 S}{d_0 + S}(\mu_0 - \hat{\mu})(\mu_0 - \hat{\mu})^T\right]\\
&= \frac{S}{\nu_0 + S}\hat{\Sigma} + \frac{\nu_0}{\nu_0 + S}\Sigma_0 + \frac{d_0 S}{(d_0 + S)(\nu_0 + S)}(\mu_0 - \hat{\mu})(\mu_0 - \hat{\mu})^T.
\end{aligned}$$

Equation (*) holds since

$$\begin{aligned}
&(d_0 + S)\left(\mu - \frac{S\hat{\mu} + d_0\mu_0}{d_0 + S}\right)^T \Sigma^{-1}\left(\mu - \frac{S\hat{\mu} + d_0\mu_0}{d_0 + S}\right)\\
&\quad + \frac{d_0 S}{d_0 + S}(\hat{\mu} - \mu_0)^T\Sigma^{-1}(\hat{\mu} - \mu_0)\\
&= \frac{1}{d_0 + S}\Big[\big((d_0 + S)\mu - S\hat{\mu} - d_0\mu_0\big)^T\Sigma^{-1}\big((d_0 + S)\mu - S\hat{\mu} - d_0\mu_0\big)\\
&\quad + d_0 S(\hat{\mu} - \mu_0)^T\Sigma^{-1}(\hat{\mu} - \mu_0)\Big]\\
&= \frac{1}{d_0 + S}\Big[\big((d_0 + S)(\mu - \hat{\mu}) + d_0(\hat{\mu} - \mu_0)\big)^T\Sigma^{-1}\big((d_0 + S)(\mu - \mu_0)\\
&\quad + S(\mu_0 - \hat{\mu})\big) + d_0 S(\hat{\mu} - \mu_0)^T\Sigma^{-1}(\hat{\mu} - \mu_0)\Big]\\
&= (d_0 + S)\,(\mu - \hat{\mu})\Sigma^{-1}(\mu - \mu_0) + S(\mu - \hat{\mu})\Sigma^{-1}(\mu_0 - \hat{\mu})\\
&\quad + d_0(\hat{\mu} - \mu_0)\Sigma^{-1}(\mu - \mu_0) + d_0 S(\hat{\mu} - \mu_0)^T\Sigma^{-1}(\mu_0 - \hat{\mu})\\
&\quad + d_0 S(\hat{\mu} - \mu_0)^T\Sigma^{-1}(\hat{\mu} - \mu_0)\\
&= (d_0 + S)(\mu - \hat{\mu})\Sigma^{-1}(\mu - \mu_0) + S(\mu - \hat{\mu})\Sigma^{-1}(\mu_0 - \hat{\mu})\\
&\quad + d_0(\hat{\mu} - \mu_0)\Sigma^{-1}(\mu - \mu_0)\\
&= d_0\left((\mu - \hat{\mu})\Sigma^{-1}(\mu - \mu_0) + (\hat{\mu} - \mu_0)\Sigma^{-1}(\mu - \mu_0)\right)\\
&\quad + S\left((\mu - \hat{\mu})\Sigma^{-1}(\mu_0 - \hat{\mu}) + (\hat{\mu} - \mu_0)\Sigma^{-1}(\mu - \mu_0)\right)\\
&= d_0(\mu - \mu_0)\Sigma^{-1}(\mu - \mu_0) + S(\mu - \hat{\mu})\Sigma^{-1}(\mu - \hat{\mu}).
\end{aligned}$$

Hence, it holds that

$$\mu, \Sigma \mid x_1, \ldots, x_S \sim \mathcal{NIW}(\mu_1, d_1, \Sigma_1, \nu_1). \tag{7.3}$$

From this posterior we can now either determine point estimates for the classical portfolio optimization or we can create an appropriate uncertainty set to use in the robust problem.

Bayesian point estimates – continuous prior

To obtain point estimates for μ and Σ, the respective marginal distributions of $\mu \mid x_1, \ldots, x_S$ and $\Sigma \mid x_1, \ldots, x_S$ are needed.

In Proposition D.15 it is shown that the marginal distribution of μ conditioned on the sample is given by a Student-t distribution:

$$\mu \mid x_1, \ldots, x_S \sim St\left(\mu_1, \frac{\nu_1}{\nu_1 - n + 1} \cdot \frac{\Sigma_1}{d_1}, \nu_1 - n + 1\right). \tag{7.4}$$

Thus, the Bayesian point estimate (see Equation (7.1)) for the return vector is given by

$$\begin{aligned} \hat{\mu}_B = \mathbf{E}[\mu \mid x_1, \ldots, x_S] &= \mu_1 \\ &= \frac{d_0}{d_0 + S}\mu_0 + \frac{S}{d_0 + S}\hat{\mu}, \end{aligned} \tag{7.5}$$

i.e. the final estimate is a convex combination of the prior and the market, the weights of each part are determined by the sample size S and the prior parameter d_0 which can as well be interpreted as a sample size where the prior return assumption is calculated from. Thus, the more data there are in the historical sample (i.e. the larger S is), the more influence the MLE of the data becomes in the final estimate compared to the prior assumption – and vice versa. Having no data at all ($S = 0$), the final return estimate is obviously simply the prior. For the sample size tending to infinity, the prior is less and less important and the posterior estimate equals the average of the data.

The marginal distribution of Σ conditioned on the sample is already explicitly given when having a normal inverse Wishart distribution:

$$\Sigma \mid x_1, \ldots, x_S \sim \mathcal{IW}(\nu_1 \Sigma_1, \nu_1 + n + 1).$$

Thus, the Bayesian point estimate for the covariance matrix is given by

$$\begin{aligned} \hat{\Sigma}_B &= \mathbf{E}[\Sigma \mid x_1, \ldots, x_S] \\ &= \frac{\nu_1 \Sigma_1}{\nu_1 + n + 1 - 2n - 2} = \frac{\nu_1}{\nu_1 - n - 1}\Sigma_1 \\ &= \frac{1}{\nu_0 + S - n - 1}\left[S\hat{\Sigma} + \nu_0 \Sigma_0 + \frac{d_0 S}{d_0 + S}(\mu_0 - \hat{\mu})(\mu_0 - \hat{\mu})^T\right] \end{aligned} \tag{7.6}$$

Remark 7.2 (Consistency of the Bayes point estimates). *This Bayesian point estimates can – like any other statistical estimator – be investigated with respect to consistency, the asymptotic behavior for $S \to \infty$. In the limit, the estimates $\hat{\mu}_B$ and $\hat{\Sigma}_B$ reduce to*

$$\hat{\mu}_B = \hat{\mu} \quad \textit{and} \quad \hat{\Sigma}_B = \hat{\Sigma},$$

and since $\hat{\mu}$ and $\hat{\Sigma}$ as the maximum likelihood estimators on a sample of size S are consistent, the Bayes estimates are consistent as well.

Remark 7.3. *As the Bayes point estimate for the return vector is given as a convex combination of the prior and the market mean, we can influence selective assets individually by setting the prior μ_0 equal to $\hat{\mu}$ for the components we do not wish to change or for which we do not have a prior assumption. Expressing "partial priors" is hence possible, and only the chosen assets are influenced by different prior values. In other words, making a prior assumption only for one asset does not affect the final return point estimate for the other assets.*

Bayesian uncertainty set – continuous prior

To create an uncertainty set for the robust portfolio optimization via Bayesian parameter estimation, we also continue from the posterior distribution of $\mu, \Sigma \mid x_1, \ldots, x_S$ as given in Equation (7.3). Instead of considering the expected values of the respective marginal distributions as point estimates, we use the *distributional* information to define an appropriate uncertainty set.

Recall the posterior distribution obtained when assuming a normal inverse Wishart distribution as prior:

$$\mu, \Sigma \mid x_1, \ldots, x_S \sim \mathcal{NIW}(\mu_1, d_1, \Sigma_1, \nu_1).$$

Since in this Bayesian approach both μ and Σ are exposed to uncertainty, i.e. are given in distributional terms, we now have different possibilities to perform a robust portfolio optimization, as analyzed in Meucci [57]:

(a) Only the return vector μ is assumed to be uncertain, the covariance matrix is supposed to be given by the Bayesian point estimate[1] $\hat{\Sigma}_B$, meaning that we do not explicitly account for uncertainty for the covariance. We thus create an uncertainty set only for the vector of expected returns. Hence, we need the marginal posterior distribution of $\mu \mid x_1, \ldots, x_S$, i.e. the unconditional posterior (unconditional with respect to Σ) which is given by a Student-t distribution as described in Equation (7.4):

$$\mu \mid x_1, \ldots, x_S \sim St\left(\mu_1, \frac{\nu_1}{\nu_1 - n + 1} \cdot \frac{\Sigma_1}{d_1}, \nu_1 - n + 1\right).$$

The (most natural) uncertainty set can now be formed by the confidence ellipsoid centered at the expectation, shaped by the covariance matrix and the size chosen according to the desired confidence, i.e.

$$\mathcal{U}_B = \left\{\mu \in \mathbb{R}^n \mid (\mu - m_B)^T (C_B)^{-1} (\mu - m_B) \leq \delta^2\right\} \tag{7.7}$$

[1] Note that when letting $\nu_0 = \infty$ and $\Sigma_0 = \hat{\Sigma}$ in the prior assumption, the Bayesian posterior simplifies to $\mu \sim \mathcal{N}\left(\mu_1, \frac{\hat{\Sigma}}{d_1}\right)$ and $\hat{\Sigma}_B = \hat{\Sigma}$.

or, respectively, expressed as a joint uncertainty set for both μ and Σ,

$$\mathcal{U}_B = \left\{\mu \in \mathbb{R}^n \mid (\mu - m_B)^T (C_B)^{-1} (\mu - m_B) \leq \delta^2\right\} \times \{\hat{\Sigma}_B\}$$

with

$$m_B = \mathbf{E}[\mu \mid x_1, \ldots, x_S] = \mu_1 = \frac{d_0}{d_0 + S}\mu_0 + \frac{S}{d_0 + S}\hat{\mu}, \tag{7.8}$$

$$C_B = \mathbf{Cov}[\mu \mid x_1, \ldots, x_S] = \frac{\nu_1}{\nu_1 - n - 1} \cdot \frac{\Sigma_1}{d_1}. \tag{7.9}$$

Note that the midpoint of the ellipsoidal uncertainty set coincides with the point estimate $\hat{\mu}_B$ as given in Equation (7.5). This is not surprising due to the unimodal and ellipsoidal structure of the Student-t distribution.

(b) A situation in practical problems which can e.g. occur when only the minimum variance portfolio is of interest, is that the return vector is assumed to be known and only the covariance matrix Σ is exposed to uncertainty. From the posterior distribution (see Equation (7.3)) we obtain that the marginal distribution of $\Sigma \mid x_1, \ldots, x_S$ is given by an inverse Wishart distribution:

$$\Sigma \mid x_1, \ldots, x_S \sim \mathcal{IW}(\nu_1 \Sigma_1, \nu_1 + n + 1).$$

As there exist closed form expressions (see e.g. Meucci [57], page 85) for the moments of a Wishart distribution, an uncertainty set can be formed by the respective confidence ellipsoid, see e.g. Meucci [57], Section 7.2, and selective parts of the calculations in Propositions 5.10 and 5.11.

(c) If both the return vector μ and the covariance matrix Σ are exposed to uncertainty, the two variables can be interpreted as one variable by combining them in the form

$$\theta := \begin{pmatrix} \mu \mid x_1, \ldots, x_S \\ \mathrm{vec}(\Sigma \mid x_1, \ldots, x_S) \end{pmatrix}.$$

For this joint variable, the first two moments can as well be calculated and an uncertainty set as given in Equation (5.5) in Section 5.2 can be defined. We have shown extensive calculations in Propositions 5.10 and 5.11 how to rewrite such a joint uncertainty set for the return vector and the covariance matrix and how to solve for the worst case parameters. For further analysis, see also Meucci [57], Section 7.2.

Remark 7.4 (Consistency of the Bayes uncertainty set). *Considering a Bayesian uncertainty set for the return vector μ as in Equation (7.7), we naturally want to study if this is a consistent uncertainty set, recall Definition 6.7. We have already seen that the Bayesian parameter estimates $\hat{\mu}_B$ and $\hat{\Sigma}_B$ are consistent, hence also $m_B = \hat{\mu}_B$ and $S \cdot C_B = \frac{S}{d_0+S}\hat{\Sigma}_B$. We thus have the same structure*

as in the definition of an uncertainty set by a confidence ellipsoid: a consistent estimator as midpoint, and the shape is given by $\frac{1}{S}$ times a consistent matrix estimate. Proceeding as in Proposition 6.8 hence gives consistency of the Bayesian uncertainty set.

Example 7.5. *This example illustrates the robust portfolio optimization if the Bayesian approach is used to define the uncertainty set for the return vector. As prior assumption we use the following parameters:*

- *$\nu_0 = S$ and $d_0 = S$, reflecting that the prior assumptions μ_0 and Σ_0 could come from a different sample of the same size.*
- *The vector μ_0 is given by the median of the underlying data, i.e. a different statistical estimator for the mean of an elliptical distribution is used.*
- *To determine the covariance prior Σ_0 we reduce the correlation between the individual assets to represent the assumption of more independency. The covariance matrix is then obtained by multiplication with the volatilities of the assets which are simply calculated from the data.*

Table 7.1 summarizes the (annualized) prior values for the time point 01.11.2003, the same time as in the investigations in Chapters 4 and 5. For comparison we also recall the market parameters at that time.

Prior	return	volatility	correlation matrix				
Lehman Eur	10.2%	3.1%	1.00	0	0	0	0
Stoxx 50	11.9%	22.1%	0	1.00	0.20	0.20	0.20
Stoxx SC	32.5%	14.6%	0	0.20	1.00	0.20	0.20
MSCI Japan	27.5%	19.5%	0	0.20	0.20	1.00	0.20
MSCI EM	47.3%	13.9%	0	0.20	0.20	0.20	1.00
Market	return	volatility	correlation matrix				
Lehman Eur	9.2%	3.1%	1.00	-0.41	-0.36	-0.09	-0.21
Stoxx 50	5.9%	22.1%	-0.41	1.00	0.80	0.29	0.60
Stoxx SC	27.0%	14.6%	-0.36	0.80	1.00	0.50	0.70
MSCI Japan	19.0%	19.5%	-0.09	0.29	0.50	1.00	0.57
MSCI EM	32.2%	13.9%	-0.21	0.60	0.70	0.57	1.00

Table 7.1: Annualized Bayesian prior assumptions and market characteristics on 01.11.2003.

Calculating the midpoint and the shape matrix according to the above formulas and using a 60% confidence to determine the size, the uncertainty set can be created. Figure 7.1 shows the projection of the ellipsoid onto the two assets Lehman Euro and Stoxx 50.

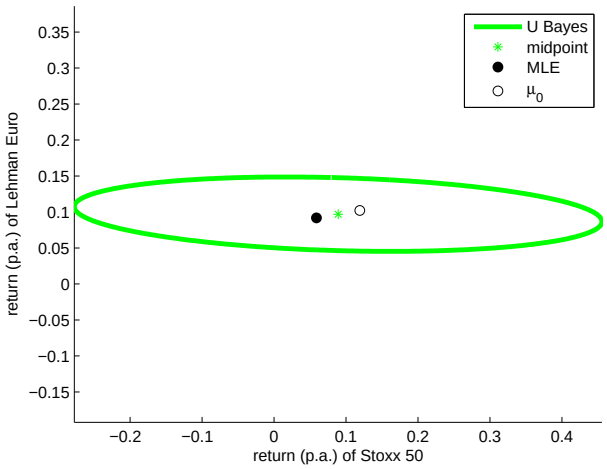


Figure 7.1: Bayes uncertainty set.

Using this uncertainty set, the optimal robust portfolio and the corresponding efficient frontier can be calculated. Figure 7.2 plots the Bayesian efficient frontier together with the classical one for comparison, and in Figure 7.3 the associated portfolio allocations are shown.

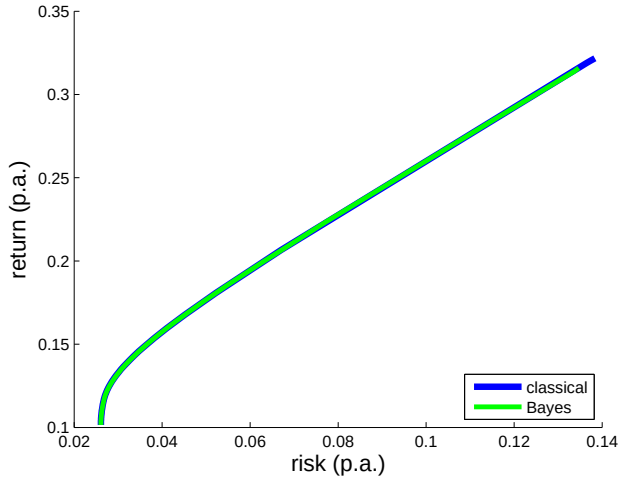


Figure 7.2: Classical and Bayesian efficient frontier on 01.11.2003.

In this sample, the Bayesian and the classical efficient frontier are very similar. From the weight plots it can be seen that the optimal portfolios are slightly different, but the resulting differences in the risk and return characteristics are not large enough to be observable in the graph. Analogous to the robustification used in Section 5.3, the Bayesian approach leads here to a shorter efficient frontier.

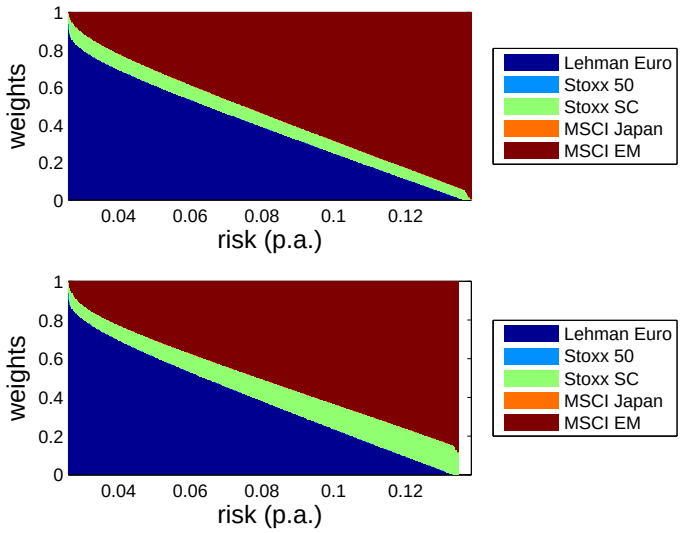
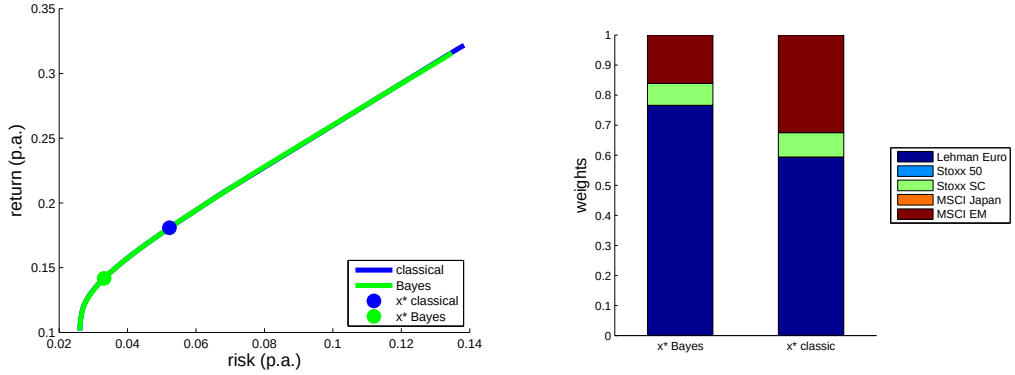


Figure 7.3: Classical and Bayesian efficient portfolios on 01.11.2003.

Finally, we again consider an investor with a given risk-aversion, expressed in terms of a fixed value for λ, and illustrate in Figure 7.4 his modified position and allocation when perfoming a robust portfolio optimization with an uncertainty set created using the Bayesian approach. As already discussed in Chapter 5, applying the robust counterpart approach generally leads to more conservative portfolios, i.e. lying closer to the minimum variance portfolio.



(a) positions on efficient frontiers (b) corresponding allocations

Figure 7.4: Implications of robust Bayes optimization for a particular investor.

To illustrate that the same Bayesian approach can also lead to rather different portfolio allocations compared to the classical setting, we pick a second point in time and perform the analogous calculations as in Example 7.5.

Example 7.6. *In this example we choose the 07.08.2004 and calculate the classical and the robust Bayesian efficient frontier and the associated optimal portfolios. For creating the uncertainty set, we again use the median of the respective data sample as prior for the return, and the covariance prior is obtained as above by using the current volatilities of the assets and the correlation matrix from Table 7.1 with the assumption of more independent assets.*

Figures 7.5 and 7.6 illustrate the efficient frontiers and the corresponding portfolio allocations.

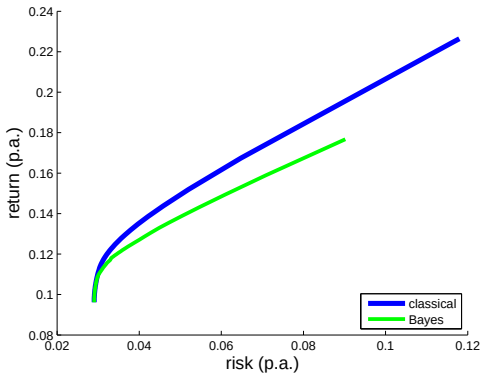


Figure 7.5: Classical and Bayesian efficient frontier on 07.08.2004.

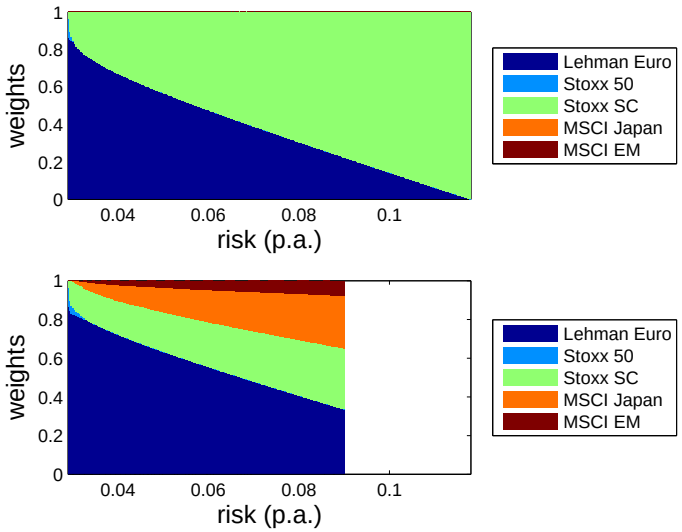


Figure 7.6: Classical and Bayesian efficient portfolios on 07.08.2004.

As can be seen, the robust Bayesian approach suggests to invest in more diversified portfolios compared to the classical allocations. Hence, the Bayesian

approach shows a similar effect as the robustification using an uncertainty set defined through estimators, see Section 5.3.

The conjecture is that the different results in these two examples are coming from the respective prior assumptions. In the first example the mean and the median (which was chosen as prior) are more alike than in the second example. This could explain the similarity to the robustification using a confidence ellipsoid or the tendency to resemble the robust approach using different estimators, respectively.

7.1.2 Bayesian approach with a discrete prior

In this section we want to investigate the Bayesian parameter estimation approach as presented in the preceding Section 7.1.1, but in this case under the assumption of a discrete prior distribution.

In particular, we will assume that N experts have published their (point) estimates for the parameters μ and Σ. As we do not consider all experts to be equally trustworthy, we assign probabilities $p_1, \ldots, p_N$ with $\sum_{j=1}^{N} p_j = 1$ to their forecasts. To avoid confusion with $\hat{\mu}_1$ or $\hat{\Sigma}_1$ from the continuous section, we denote the experts' estimates by r_j and C_j, respectively, $j = 1, \ldots, N$. Mathematically, we thus assume the following discrete prior distribution:

$$(\mu, \Sigma) = \begin{cases} (r_1, C_1) & \text{with probability } p_1 \\ \quad \vdots & \\ (r_N, C_N) & \text{with probability } p_N \end{cases}$$

with

$$\sum_{j=1}^{N} p_j = 1.$$

In terms of a density function, this can be written as

$$\varphi_{\text{prior}}(\mu, \Sigma) = \sum_{j=1}^{N} p_j \, \delta_{r_j}(\mu) \, \delta_{C_j}(\Sigma).$$

with δ_{z_0} denoting the density of the Dirac measure, i.e. the entire mass is concentrated at the point z_0:

$$\int_A \delta_{z_0}(z) dz = \begin{cases} 1 & \text{if } z_0 \in A \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the (prior) point estimates which could be used in the classical portfolio optimization are again calculated by minimizing the loss function and are hence

given by

$$\hat{\mu}_{\text{discrete}} = \mathbf{E}[\mu] = \sum_{j=1}^{N} p_j r_j,$$
$$\hat{\Sigma}_{\text{discrete}} = \mathbf{E}[\Sigma] = \sum_{j=1}^{N} p_j C_j$$

according to Equation (7.1).

Remark 7.7. *Instead of choosing as estimates the ones minimizing the expected loss, we could as well use the maximum likelihood estimators, simply given by the pair of parameters with the highest probability p_j.*

We now want to calculate Bayesian parameter estimates incorporating both the prior distribution and the market information which is given in terms of a data sample. Analogous to the continuous case, we distinguish the individual steps in the calculations.

1. *Prior distribution*
   As prior distribution we assume the discrete distribution from above, i.e.

$$\varphi_{\text{prior}}(\mu, \Sigma) = \sum_{j=1}^{N} p_j \, \delta_{r_j}(\mu) \, \delta_{C_j}(\Sigma).$$

2. *Market information*
   The market information is – as in the previous section – given by the density function of the sample where it still holds that $X_i \mid \mu, \Sigma \sim \mathcal{N}(\mu, \Sigma)$ i.i.d., $i = 1, \ldots, S$, i.e. the sample is normally distributed with the unknown parameters μ and Σ. Recall Formula (7.2):

$$\begin{aligned}\varphi(x_1, \ldots, x_S \mid \mu, \Sigma) &= \prod_{s=1}^{S} \varphi(x_s \mid \mu, \Sigma) \\ &= \left(\frac{1}{(2\pi)^{\frac{n}{2}}}\right)^S |\Sigma|^{-\frac{S}{2}} \exp\left\{-\frac{1}{2}[S(\hat{\mu}-\mu)^T\Sigma^{-1}(\hat{\mu}-\mu) + \operatorname{tr}(S\hat{\Sigma}\Sigma^{-1})]\right\}.\end{aligned}$$

   Defining for notational convenience

$$z(\mu, \Sigma) := S(\hat{\mu}-\mu)^T\Sigma^{-1}(\hat{\mu}-\mu) + \operatorname{tr}(S\hat{\Sigma}\Sigma^{-1}),$$

   the joint density of the sample simplifies to

$$\varphi(x_1, \ldots, x_S \mid \mu, \Sigma) = \left(\frac{1}{(2\pi)^{\frac{n}{2}}}\right)^S |\Sigma|^{-\frac{S}{2}} \exp\left\{-\frac{1}{2}z(\mu, \Sigma)\right\}.$$

3. *Posterior distribution*
The posterior distribution of $\mu, \Sigma \mid x_1, \ldots, x_S$ is calculated analogously to the continuous case.

$$\begin{aligned}
\varphi_{\text{post}}(\mu, \Sigma \mid x_1, \ldots, x_S) &= \gamma \varphi(x_1, \ldots, x_S \mid \mu, \Sigma)\, \varphi_{\text{prior}}(\mu, \Sigma) \\
&= \gamma \left(\frac{1}{(2\pi)^{\frac{n}{2}}} \right)^S |\Sigma|^{-\frac{S}{2}} \exp\left\{ -\frac{1}{2} z(\mu, \Sigma) \right\} \\
&\quad \cdot \sum_{j=1}^{N} p_j\, \delta_{r_j}(\mu)\, \delta_{C_j}(\Sigma) \\
&= \sum_{j=1}^{N} \underbrace{\gamma \left(\frac{1}{(2\pi)^{\frac{n}{2}}} \right)^S |\Sigma|^{-\frac{S}{2}} \exp\{-\frac{1}{2} z(\mu, \Sigma)\} p_j}_{=:\tilde{p}_j}\, \delta_{r_j}(\mu)\, \delta_{C_j}(\Sigma) \\
&= \sum_{j=1}^{N} \tilde{p}_j\, \delta_{r_j}(\mu)\, \delta_{C_j}(\Sigma)
\end{aligned}$$

where $\tilde{p}_j, j = 1, \ldots, N$ (or γ respectively) are chosen appropriately such that $\sum_{j=1}^{N} \tilde{p}_j = 1$ again, i.e.

$$\begin{aligned}
& \sum_{j=1}^{N} \tilde{p}_j = 1 \\
\Leftrightarrow \quad & \sum_{j=1}^{N} \gamma \left(\frac{1}{(2\pi)^{\frac{n}{2}}} \right)^S |\Sigma|^{-\frac{S}{2}} \exp\{-\frac{1}{2} z(r_j, C_j)\} p_j = 1 \\
\Leftrightarrow \quad & \gamma \left(\frac{1}{(2\pi)^{\frac{n}{2}}} \right)^S |\Sigma|^{-\frac{S}{2}} \sum_{j=1}^{N} \exp\{-\frac{1}{2} z(r_j, C_j)\} p_j = 1,
\end{aligned}$$

hence

$$\gamma = \frac{(2\pi)^{\frac{nS}{2}} |\Sigma|^{\frac{S}{2}}}{\sum_{j=1}^{N} \exp\{-\frac{1}{2} z(r_j, C_j)\} p_j}.$$

Thus, we finally have

$$\tilde{p}_j = \frac{\exp\{-\frac{1}{2} z(r_j, C_j)\} p_j}{\sum_{j=1}^{N} \exp\{-\frac{1}{2} z(r_j, C_j)\} p_j}.$$

From these calculations it can be seen that the posterior distribution is again a discrete distribution, just the probabilities for the N outcomes are changed when moving from the prior to the posterior. Hence, whenever assuming a discrete prior

distribution, the posterior remains discrete and does not change to a continuous distribution, even if a continuous distribution is imposed via the market condition.

Accordingly, the Bayesian point estimates for the parameters under the assumption of a discrete prior distribution are given by

$$\hat{\mu}_{B,\text{discrete}} = \mathbf{E}[\mu \mid x_1, \ldots, x_S] = \sum_{j=1}^{N} \tilde{p}_j \, r_j,$$

$$\hat{\Sigma}_{B,\text{discrete}} = \mathbf{E}[\Sigma \mid x_1, \ldots, x_S] = \sum_{j=1}^{N} \tilde{p}_j \, C_j.$$

Here, we could again choose as well the maximum likelihood estimators, i.e. the pair with the largest posterior probability $\tilde{p}_j$ instead of the expectations which represent the estimators minimizing the expected loss.

We now also want to define an uncertainty set only around the return vector μ for the robust portfolio optimization based on the result of the Bayesian calculations. In analogy to the continuous Bayesian approach, a confidence uncertainty set can be built. We have finitely many expert opinions, weighted with different posterior probabilities $\tilde{p}_j$. To obtain an uncertainty set to the confidence level α, we first reorder the parameter pairs such that their associated probabilities are in descending order, i.e. such that $\tilde{p}_{(1)} \geq \ldots \geq \tilde{p}_{(N)}$. Then we determine a number $l \in \mathbb{N}, l \leq N$ such that

$$\sum_{j=1}^{l} \tilde{p}_j \geq \alpha \quad \text{and} \quad \sum_{j=1}^{l-1} \tilde{p}_j < \alpha$$

and use the respective points $r_{(1)}, \ldots, r_{(l)}$ to create an uncertainty set for the return. This is naturally given by the convex hull of these points, i.e.

$$\mathcal{U}_{\text{discrete, conv}} = \text{conv}\left(\left\{r_{(1)}, \ldots, r_{(l)}\right\}\right).$$

As we prefer using ellipsoids as uncertainty sets, we can approximate the confidence set by any of the previously described methods from Section 5.3 to define an ellipsoid containing a number of given points. The straightforward approach is again to find the smallest ellipsoid centered at the expectation and shaped by the covariance matrix based on the l points that are considered. Hence, we first calculate the relative probabilities for those l vectors by

$$\tilde{\tilde{p}}_{(j)} := \frac{\tilde{p}_{(j)}}{\sum_{k=1}^{l} \tilde{p}_{(k)}}.$$

Then, the midpoint of the ellipsoid is determined by

$$m_{\text{disc}} = \mathbf{E}[\mu] = \sum_{j=1}^{l} \tilde{\tilde{p}}_{(j)} r_{(j)}$$

and the shape matrix calculates to

$$\begin{aligned} C_{\text{disc}} &= \mathbf{Cov}[\mu] = \mathbf{E}[\mu\mu^T] - \mathbf{E}[\mu]\,\mathbf{E}[\mu]^T \\ &= \sum_{j=1}^{l} \tilde{\tilde{p}}_{(j)} r_{(j)} r_{(j)}^T - \left(\sum_{j=1}^{l} \tilde{\tilde{p}}_{(j)} r_{(j)}\right)\left(\sum_{k=1}^{l} \tilde{\tilde{p}}_{(k)} r_{(k)}\right)^T \\ &= \sum_{j=1}^{l} \tilde{\tilde{p}}_{(j)}(1-\tilde{\tilde{p}}_{(j)}) r_{(j)} r_{(j)}^T - \sum_{j=1}^{l} \sum_{\substack{k=1 \\ k\neq j}}^{l} \tilde{\tilde{p}}_{(j)} \tilde{\tilde{p}}_{(k)} r_{(j)} r_{(k)}^T. \end{aligned}$$

Note that for this shape matrix to be invertible it has to hold that $l \geq n$. The uncertainty set can then be described by

$$\mathcal{U}_{\text{discrete}} = \left\{\mu \in \mathbb{R}^n \mid (\mu - m_{\text{disc}})^T C_{\text{disc}}^{-1} (\mu - m_{\text{disc}}) \leq \delta^2\right\} \times \{\hat{\Sigma}_{B,\text{discrete}}\}$$

with the size of the ellipsoid chosen the smallest possible such that $r_{(1)}, \ldots, r_{(l)}$ are lying within, i.e.

$$\delta^2 = \max_{j=1,\ldots,l} (r_{(j)} - m_{\text{disc}})^T C_{\text{disc}}^{-1} (r_{(j)} - m_{\text{disc}}).$$

Alternatively, we could determine the minimum volume ellipsoid containing the desired points.

To round up this section, we also illustrate the discrete robust Bayesian approach in the same example as used before.

Example 7.8. *For the discrete Bayesian approach several experts' opinions are necessary to form a prior assumption. In view of the definition of an uncertainty set where we will need more estimates than assets, we start with the following 8 discrete prior values for the return:*

- *The five different point estmates as given in Section 4.3, i.e. MLE, median, quartile estimator, Huber estimator and the trimmed mean.*
- *Additionally we use three long term estimators that are calculated based on the entire historical data sample. They thus consider not only the last year's performance of the assets, but a longer average. Here we choose the mean, the median and as a more robust estimator the trimmed mean as long term estimators.*

We assume furthermore that the experts do not have a particular opinion about the covariance matrix, hence we use $C_j = \hat{\Sigma}$ for $j = 1, \ldots, 8$. The probabilities assigned to the various discrete priors are supposed to be given by the vector $p = (20\%, 20\%, 10\%, 10\%, 10\%, 10\%, 10\%, 10\%)^T$. With this prior setting we obtain at the time 01.11.2003 the following vector of posterior probabilities (rounded to %) after taking the market into account:

$$\tilde{p} = (40\%, 17\%, 3\%, 15\%, 19\%, 1\%, 3\%, 2\%)^T.$$

It was expected that the posterior probabilities of the long term estimators are reduced since these estimates are rather different from the market represented by the current data sample. To create an uncertainty set, we need to sort this vector in descending order and take the first l such that $\sum_{j=1}^{l} \tilde{p}_j \geq 60\%$. The first (MLE) and the fifth (trimmed mean) estimators together already almost suffice, but we nevertheless take the largest 6 out of the 8 estimators for assuring positive definiteness of the needed shape matrix. Note that this corresponds to a confidence of roughly 97%. Compared to the uncertainty set defined using the five different estimators, we here exchange the quartile estimator with the long term median and additionally use the long term trimmed mean.

With the respective estimators we hence determine the midpoint, the shape and the size of the uncertainty set as given above. The result in shown in Figure 7.7.

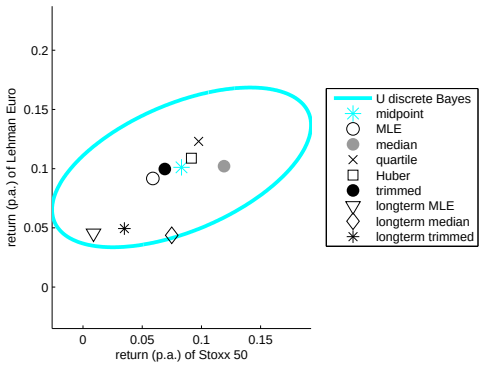


Figure 7.7: Discrete Bayes uncertainty set.

Analogous to the previous examples, we illustrate the discrete Bayesian efficient frontier compared to the classical one in Figure 7.8 and the according optimal portfolio allocations in Figure 7.9.

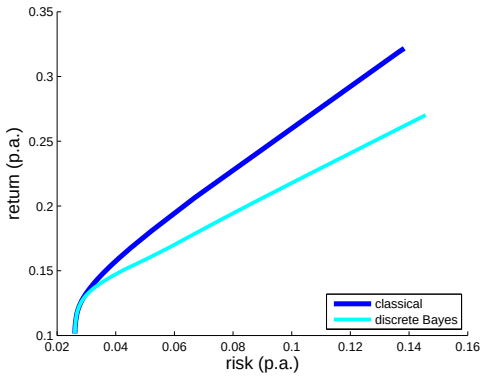


Figure 7.8: Classical and discrete Bayesian efficient frontier on 01.11.2003.

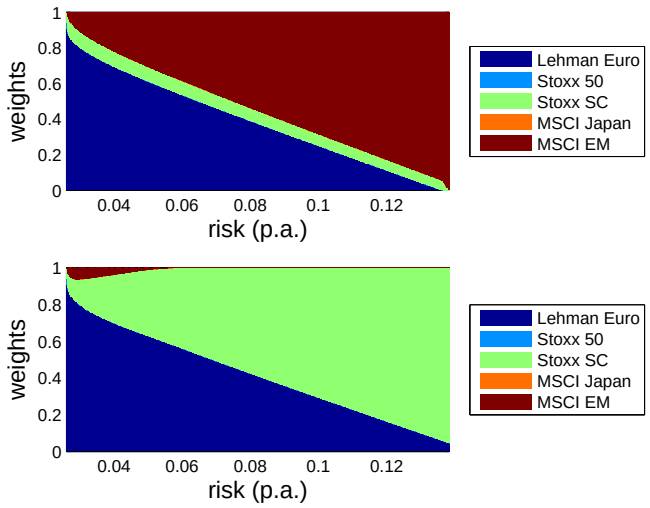
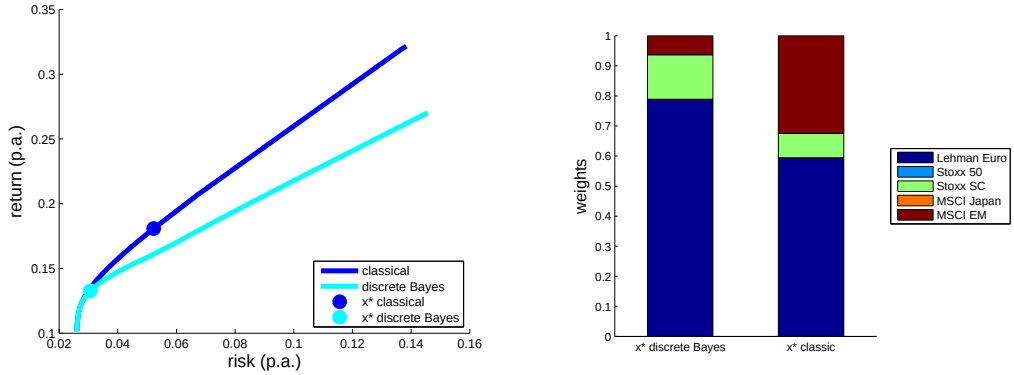


Figure 7.9: Classical and discrete Bayesian efficient portfolios on 01.11.2003.

The tendency of the efficient frontier and especially the optimal allocations show the similarity to the robust portfolio optimization from Section 5.3. This was expected since (almost) the same points were used to create the uncertainty sets for the return, only applying different methods.

To illustrate the effects of such a discrete Bayesian robustification, Figure 7.10 again shows the position on the efficient frontiers and the corresponding portfolio allocations.



(a) positions on efficient frontiers (b) corresponding allocations

Figure 7.10: Implications of discrete robust Bayes optimization for a particular investor.

7.2 Black-Litterman approach

A different approach combining market information and external knowledge is the *Black-Litterman* model, see Black and Litterman [15], which is, to the best of our knowledge, so far only used to obtain a point estimate for the return vector for the classical portfolio optimization problem. We will illustrate that it can naturally be used for the definition of an uncertainty set as well since the distribution of the posterior estimate is known.

The market information is as in the Bayes model given by a sample of historical data. The external knowledge in the Bayes model was described in terms of a distributional assumption about the parameters, in the Black-Litterman model it is incorporated via individual opinions – the so-called "forecasts" – about selective assets.

In the Black-Litterman setting, only the return vector μ is explicitly exposed to uncertainty, the covariance matrix Σ is supposed to be known, i.e. estimated by $\hat{\Sigma}$ without uncertainty.

The *prior assumption* in this framework is given by the following distribution:

$$\mu \sim \mathcal{N}(\hat{\mu}, \tau\hat{\Sigma})$$

with $\hat{\mu}$ and $\hat{\Sigma}$ being the maximum likelihood estimators based on the data realizations $x_1, \ldots, x_S$ and $\tau \in [0, 1]$ representing the confidence in this prior estimate.

Since the covariance matrix is supposed to be fixed, investor forecasts can only be made for the uncertain vector μ. Those forecasts are not only a simple point estimate but some absolute or relative opinions about the return vector, mathematically expressed in the form

$$Q = P\mu + \varepsilon$$

with $Q \in \mathbb{R}^m$, $P \in \mathbb{R}^{m\times n}$, $\operatorname{rank} P = m$, $\varepsilon \sim \mathcal{N}(0, \Omega)$ and $\mathbf{Cov}[\varepsilon, \mu] = 0$. The vector Q contains the forecasted values, and the matrix P contains the information about the assets that are affected by the respective forecasts. For example, having three assets named A, B and C, and forecasting "A outperforms B by 4%" and "C has a return of 6%", Q and P would be given by

$$P = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{and} \quad Q = \begin{pmatrix} 4\% \\ 6\% \end{pmatrix}.$$

As the matrix Ω describes the variance of Q, it expresses the confidence about the individual forecasts. Combining the forecast with the prior distribution of μ, we thus obtain the conditional distribution of Q given μ as

$$Q \mid \mu \sim \mathcal{N}(P\mu, \Omega).$$

From $Q \mid \mu \sim \mathcal{N}(P\mu, \Omega)$ and $\mu \sim \mathcal{N}(\hat{\mu}, \tau\hat{\Sigma})$ we can calculate the marginal distribution of Q, the joint distribution of μ and Q and finally the conditional

distribution of μ given $Q = q$. These results are summarized in the subsequent proposition.

Proposition 7.9. *Let $\mu \sim \mathcal{N}(\hat{\mu}, \tau\hat{\Sigma})$ and let $Q = P\mu + \varepsilon$ with $Q \in \mathbb{R}^m$, $P \in \mathbb{R}^{m\times n}$, rank $P = m$, $\varepsilon \sim \mathcal{N}(0, \Omega)$ and $\mathbf{Cov}[\varepsilon, \mu] = 0$. Then, the following statements hold:*

(i) The marginal distribution of Q is given by

$$Q \sim \mathcal{N}_m(P\hat{\mu}, \tau \cdot P\hat{\Sigma}P^T + \Omega).$$

(ii) The joint distribution of μ and Q is given by

$$\begin{pmatrix} \mu \\ Q \end{pmatrix} \sim \mathcal{N}_{n+m}\left(\begin{pmatrix} \hat{\mu} \\ P\hat{\mu} \end{pmatrix}, \begin{pmatrix} \tau\hat{\Sigma} & \tau\hat{\Sigma}P^T \\ \tau P\hat{\Sigma} & \tau \cdot P\hat{\Sigma}P^T + \Omega \end{pmatrix}\right).$$

(iii) The conditional distribution $\mu \mid Q$ is given by

$$\begin{aligned} \mu \mid Q = q \sim \mathcal{N}_n\big(&\hat{\mu} + \tau\hat{\Sigma}P^T(\tau \cdot P\hat{\Sigma}P^T + \Omega)^{-1}(q - P\hat{\mu}), \\ &\tau\hat{\Sigma} - \tau\hat{\Sigma}P^T(\tau \cdot P\hat{\Sigma}P^T + \Omega)^{-1}\tau P\hat{\Sigma}\big). \end{aligned} \tag{7.10}$$

Proof.

(i) Expressing $\mu \sim \mathcal{N}(\hat{\mu}, \tau\hat{\Sigma})$ in the form $\mu = \hat{\mu} + \nu$ with $\nu \sim \mathcal{N}(0, \tau\hat{\Sigma})$ and ν independent from ε, we obtain

$$Q = P\mu + \varepsilon \;=\; P\hat{\mu} + P\nu + \varepsilon$$

which thus follows again a normal distribution, and the moments are

$$\begin{aligned} \mathbf{E}[Q] &= P\hat{\mu}, \\ \mathbf{Cov}[Q] &= \mathbf{Cov}[P\nu + \varepsilon] \\ &= P\,\mathbf{Cov}[\nu]P^T + 2P\,\mathbf{Cov}[\nu\varepsilon] + \mathbf{Cov}[\varepsilon] \\ &= \tau P\hat{\Sigma}P^T + \Omega. \end{aligned}$$

(ii) The covariance between μ and Q can be calculated as

$$\begin{aligned} \mathbf{Cov}[\mu, Q] &= \mathbf{E}[(\mu - \mathbf{E}[\mu])(Q - \mathbf{E}[Q])^T] \\ &= \mathbf{E}[(\mu - \hat{\mu})(Q - P\hat{\mu})^T] \\ &= \mathbf{E}[((\hat{\mu} + \nu) - \hat{\mu})((P\hat{\mu} + P\nu + \varepsilon) - P\hat{\mu})^T] \\ &= \mathbf{E}[\nu(P\nu + \varepsilon)^T] \\ &= \mathbf{E}[\nu\nu^T P^T] + \mathbf{E}[\nu\varepsilon^T] \\ &= \mathbf{Cov}[\nu]P^T + \mathbf{Cov}[\nu, \varepsilon] \\ &= \tau\hat{\Sigma}P^T. \end{aligned}$$

Thus, the above joint distribution follows.

(iii) This follows immediately from the conditioning formula for the normal distribution, given by Equation (D.1). □

This conditional distribution $\mu \mid Q$ can be interpreted as the posterior distribution of μ after incorporating additional information – here given in the form of experts' opinions. (Recall that in the Bayesian framework, the additional information came from the historical data and the prior assumption was e.g. given by an expert.)

For actually determining point estimates or an uncertainty set in the Black-Litterman setting, the matrix Ω expressing the confidence in the individual forecasts needs to be specified. It can be distinguished between the two cases of dependent and independent forecasts.

- In the case of independent forecasts – which we will not pursue any further – the matrix Ω is chosen as a diagonal matrix, describing a possibly different confidence for each individual forecast, i.e.

$$\Omega := \operatorname{diag}(\omega_1, \dots, \omega_m).$$

- In the case of dependent forecasts, we assume the dependence structure expressed by the original covariance matrix, which naturally has to be modified by the transition matrix P to match the structure and the dimension of the individual forecasts. By the scalar $1 - \tau$ a general confidence in the forecasts is defined.

$$\Omega := (1 - \tau) P \hat{\Sigma} P^T.$$

  Recall that the prior assumption of the market is given by

$$\mu \sim \mathcal{N}(\hat{\mu}, \tau \hat{\Sigma}).$$

  Hence, τ represent a trade-off between the market and the forecasts, with the limits $\tau = 0$ expressing complete confidence in the market and $\tau = 1$ neglecting the market and relying only on the forecasts. This is as well reflected in Equation (7.10) where e.g. for $\tau = 0$ the posterior distribution is again reduced to the market assumption. Note that since Ω is included in the formula through its inverse, a larger value of τ – hence smaller entries in Ω – represents a larger influence of the forecasts in the final outcome.

Assumption 7.10. *Throughout the rest of the chapter, we assume dependent forecasts, i.e. the confidence matrix is given by*

$$\Omega = (1 - \tau) P \hat{\Sigma} P^T.$$

Note that using this assumption, the posterior distribution $\mu \mid Q$ simplifies to

$$\begin{aligned} \mu \mid Q = q \sim \mathcal{N}_n\big(&\hat{\mu} + \tau \hat{\Sigma} P^T (P \hat{\Sigma} P^T)^{-1} (q - P\hat{\mu}), \\ &\tau \hat{\Sigma} - \tau \hat{\Sigma} P^T (P \hat{\Sigma} P^T)^{-1} \tau P \hat{\Sigma}\big). \end{aligned} \tag{7.11}$$

From the conditional or posterior distribution stated in Proposition 7.9 we can – analogous to the Bayesian setting – either determine a point estimate[2] for μ to input into the classical optimization problem or use the distribution (resp. the first two moments) to define an uncertainty set around μ for the robust optimization problem.

7.2.1 Black-Litterman point estimates

As point estimates we get for μ the expectation of the distribution given in Equation (7.11), and as no uncertainty was assumed for the covariance matrix, we simply use the MLE there:

$$\begin{aligned}\hat{\mu}_{BL} &= \mathbf{E}[\mu|Q=q] \\ &= \hat{\mu} + \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}(q - P\hat{\mu}) \qquad (7.12)\\ \hat{\Sigma}_{BL} &= \hat{\Sigma}. \qquad (7.13)\end{aligned}$$

Thus, the (posterior) Black-Litterman point estimate starts from the data estimate and modifies the respective entries according to given forecasts.

Example 7.11. *In a first example we assume that we only make absolute forecasts and one forecast for each asset, i.e. we give an opinion in form of a vector $q \in \mathbb{R}^n$, and the matrix P is then given by the identity, $P = I$.*

With these assumptions, Equation (7.12) simplifies to

$$\begin{aligned}\hat{\mu}_{BL} &= \hat{\mu} + \tau\hat{\Sigma}(\hat{\Sigma})^{-1}(q - \hat{\mu}) \\ &= \hat{\mu} + \tau(q - \hat{\mu}) \\ &= \tau q + (1-\tau)\hat{\mu},\end{aligned}$$

i.e. a convex combination of the data mean $\hat{\mu}$ and the forecasted values q which directly shows the influence of the parameter τ.

Let the MLE[3] of a data sample with 5 assets be given by

$$\hat{\mu} = (9.2\%, 5.9\%, 27.0\%, 19.0\%, 32.2\%)^T$$

and assume the forecasted (absolute) values to be

$$q = (6\%, 10\%, 17\%, 11\%, 15\%)^T.$$

With $\tau = 0.5$, the posterior Black-Litterman forecast is given by averaging both vectors, hence

$$\hat{\mu}_{BL} = (7.6\%, 8.0\%, 22.0\%, 15.0\%, 23.6\%)^T.$$

[2]Note that Σ was already assumed to be estimated sufficiently accurate by $\hat{\Sigma}$.

[3]Note that we present annualized values for comparability.

Example 7.12. *This second example illustrates that an individual view on only one particular asset influences the point estimate for all the other assets as well. We consider again*

$$\hat{\mu} = (9.2\%, 5.9\%, 27.0\%, 19.0\%, 32.2\%)^T$$

but this time the only forecast we make is that the second asset has an average performance of 10%*. Such a forecast is represented by*

$$P = (0, 1, 0, 0, 0) \quad \textit{and} \quad q = 10\%.$$

The confidence in this forecast is again supposed to be expressed by the market structure, thus in this particular case we have $\Omega = (1-\tau)\hat{\Sigma}_{2,2}$*. Using* $\tau = 0.5$*, the final estimate* $\hat{\mu}_{BL}$ *is calculated to*

$$\hat{\mu}_{BL} = (9.1\%, 8.0\%, 28.1\%, 19.6\%, 32.9\%)^T.$$

Thus, this simple example shows that a forecast on one asset influences all the other components in the final estimate as well – in contrast to the Bayes model, recall Remark 7.3.

Remark 7.13 (Consistency of the Black-Litterman point estimates). *The point estimate for* Σ *is simply given by* $\hat{\Sigma}_{BL} = \hat{\Sigma}$*, the maximum likelihood estimator of the underlying data sample, which is already known to be a consistent estimator.*

To investigate consistency of $\hat{\mu}_{BL}$*, we have to define* τ *sensibly in terms of the sample size* S*. As a larger sample of historical data suggests more reliability in the market assumption and should hence reflect an increasing influence of the market compared to the forecasts, a natural definition is* $\tau = \frac{1}{S}$*. Thus, for* S *tending to infinity, we obtain for the Black-Litterman estimate*

$$\hat{\mu}_{BL} = \hat{\mu} + \frac{1}{S}\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}(q - P\hat{\mu}) \to \hat{\mu},$$

hence it is a consistent estimator.

7.2.2 Black-Litterman uncertainty set

Besides determining Black-Litterman point estimates needed in the classical optimization problem, we can also create an uncertainty set for the parameter μ. Such an uncertainty set taking into account both the market information and the investor forecasts is described by the confidence ellipsoid characterized by the first two moments of the posterior distribution given in Equation (7.10):

$$\mathcal{U}_{BL} = \left\{\mu \in \mathbb{R}^n \mid (\mu - m_{BL})^T (C_{BL})^{-1} (\mu - m_{BL}) \leq \delta^2\right\} \tag{7.14}$$

with

$$\begin{aligned} m_{BL} &= \mathbf{E}[\mu|Q=q] \\ &= \hat{\mu} + \tau\hat{\Sigma}P^T(\tau \cdot P\hat{\Sigma}P^T + \Omega)^{-1}(q - P\hat{\mu}), \qquad (7.15) \\ &= \hat{\mu} + \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}(q - P\hat{\mu}), \\ C_{BL} &= \mathbf{Cov}[\mu|Q=q] \\ &= \tau\hat{\Sigma} - \tau\hat{\Sigma}P^T(\tau \cdot P\hat{\Sigma}P^T + \Omega)^{-1}\tau P\hat{\Sigma} \qquad (7.16) \\ &= \tau\hat{\Sigma} - \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}\tau P\hat{\Sigma}. \end{aligned}$$

Remark 7.14 (Consistency of the Black-Litterman uncertainty set). *We have already seen that under the assumption $\tau = \frac{1}{S}$ the point estimate $\hat{\mu}_{BL}$ and thus the midpoint of the confidence ellipsoid $m_{BL} = \hat{\mu}_{BL}$ is consistent. The shape of the uncertainty set is determined by*

$$\begin{aligned} C_{BL} &= \tau\hat{\Sigma} - \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}\tau P\hat{\Sigma} \\ &= \frac{1}{S}\left[\hat{\Sigma} - \hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}\frac{1}{S}P\hat{\Sigma}\right]. \end{aligned}$$

From this formula it can be observed that for $S \to \infty$ it holds

$$C_{BL} \to 0.$$

Using – as in the proof of Proposition 6.8 – the equivalent formulation for an ellipsoidal uncertainty set,

$$\mathcal{U}_{BL} = \{\mu \in \mathbb{R}^n \mid \mu = m_{BL} + \delta\, C_{BL}^{\frac{1}{2}} z, \|z\|_2 \leq 1\}$$

we straightforwardly obtain that $\mathcal{U}_{BL}$ reduces to the single point $\hat{\mu}$ eventually, since δ is fixed, i.e. bounded, and $C_{BL} \to 0$. Hence, $\mathcal{U}_{BL}$ is a consistent uncertainty set.

Example 7.15. *We apply the Black-Litterman approach to create an uncertainty set for the return with the following assumptions and forecasts:*

- *We let $\tau = \frac{1}{S}$ in the prior market model $\mu \sim \mathcal{N}(\hat{\mu}, \tau\hat{\Sigma})$ to make it comparable to the usual setting where the variance of the maximum likelihood estimator of a normally distributed sample is scaled by the sample size. Note that $\tau = 1$ would imply the estimator's variance to be of the same magnitude as the individual return data.*

- *Similar to the Bayesian case, we use the median as external opinions, hence we let $P = I$ and define q to be the median. With these assumptions, the formulas for the characteristics of the uncertainty set simplify to*

$$\begin{aligned} m_{BL} &= \hat{\mu} + \tau(q - \hat{\mu}), \\ C_{BL} &= \tau(1-\tau)\hat{\Sigma}. \end{aligned}$$

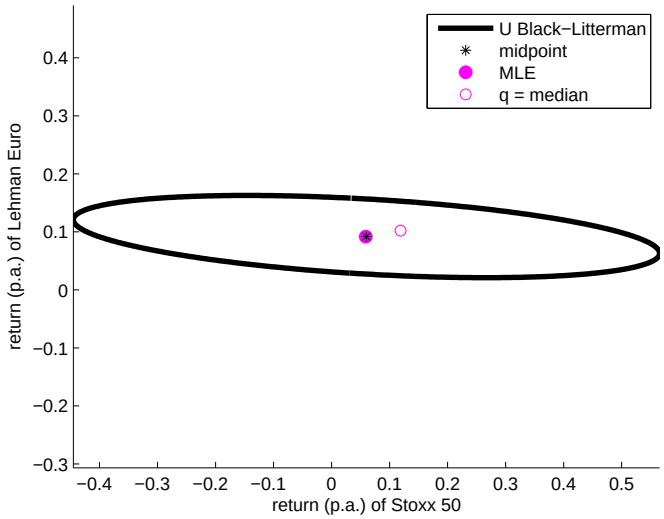


Figure 7.11: Black-Litterman uncertainty set.

The uncertainty set then looks as in Figure 7.11 at the time 01.11.2003.

The resulting plots of the efficient frontiers and the associated portfolio allocations are shown in Figures 7.12 and 7.13. Figure 7.14 finally illustrates the results for the particular investor.

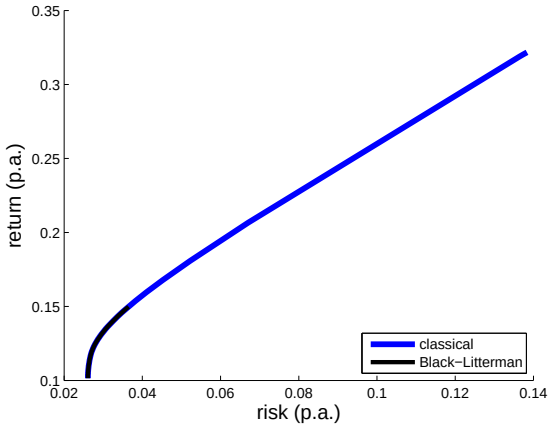


Figure 7.12: Classical and Black-Litterman efficient frontier on 01.11.2003.

Even though the efficient frontiers and the corresponding weights seem to be identical, there are very slight differences, but too small to recognize in the graphics. But the fact that the portfolios and the efficient frontiers are very similar is not surprising, since for small τ the midpoint is roughly given by the MLE $\hat{\mu}$ and the shape matrix is approximately $\frac{1}{S}\hat{\Sigma}$. This is the setting of determining an

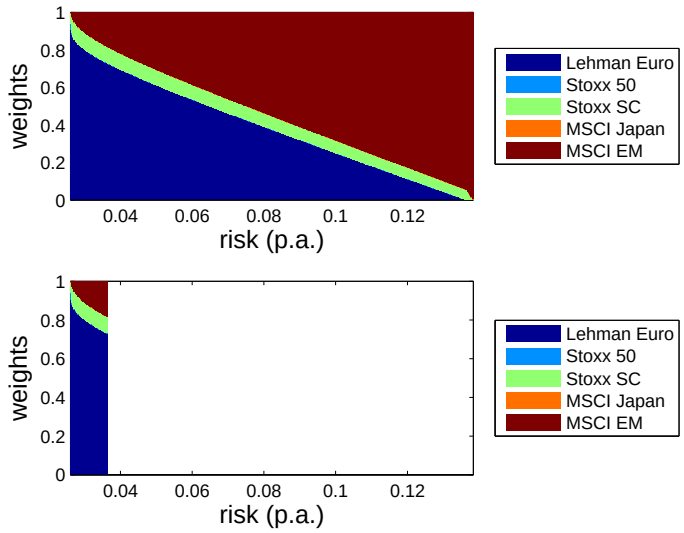
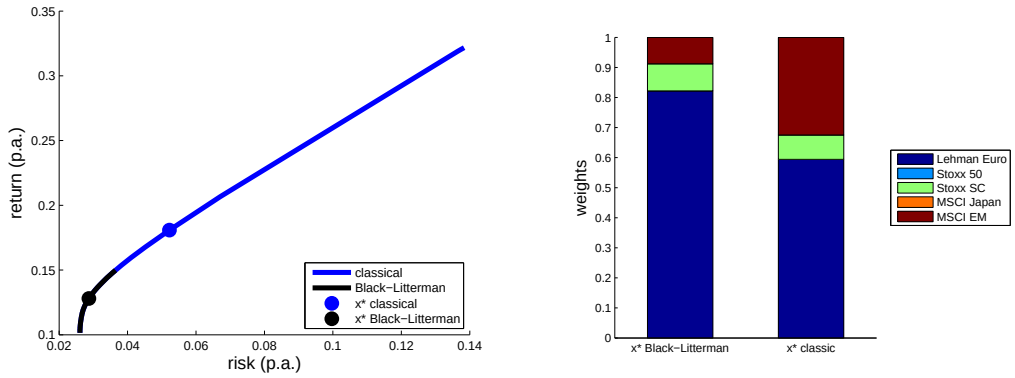


Figure 7.13: Classical and Black-Litterman efficient portfolios on 01.11.2003.



(a) positions on efficient frontiers (b) corresponding allocations

Figure 7.14: Implications of robust Black-Litterman optimization for a particular investor.

uncertainty set by a confidence ellipsoid around the MLE, as discussed in Section 5.2.1. Note that in the more general Black-Litterman setting with $P \neq I$, we obtain a similar qualitative statement. For τ close to zero, the midpoint is approximately the MLE, and for the shape matrix it holds

$$C_{BL} = \tau\left[\hat{\Sigma} - \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}P\hat{\Sigma}\right] \approx \tau\hat{\Sigma}$$

since the expression $\tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}P\hat{\Sigma}$ becomes small and is hence neglectable.

In all the previous examples we have always applied only the respective method for creating an uncertainty set and compared the robust results to the

classical setting. The following example recalls the uncertainty sets and the efficient frontiers for all the different robustifications and illustrates them simultaneously.

Example 7.16. *As the basis we consider the classical portfolio optimization problem using the maximum likelihood estimators as point estimates. For the robust portfolio optimization we recall the following approaches for creating an uncertainty set for the return vector:*

- *Using different statistical estimators, see Section 5.3.*
- *The continuous Bayesian approach, see Section 7.1.1.*
- *The discrete Bayesian approach, see Section 7.1.2.*
- *The Black-Litterman approach, see Section 7.2.*

As prior assumptions or expert opinions in the respective methods we use the values presented in the corresponding examples.

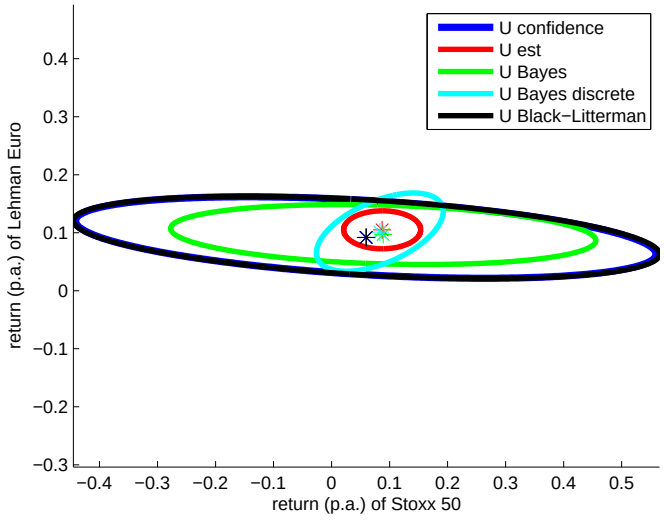


Figure 7.15: Comparison of the various uncertainty set.

The uncertainty sets (i.e. the projection onto 2 selected assets) are shown in Figure 7.15. As discussed before, the uncertainty sets in the Bayesian and the Black-Litterman approach resemble the confidence ellipsoid for the MLE which is also reflected below in Figures 7.16 and 7.17 illustrating the efficient frontiers and the portfolio allocations. The uncertainty sets in the case of various estimators and in the discrete Bayes model are shaped differently and hence also result in a modified efficient frontier and changed portfolio allocations.

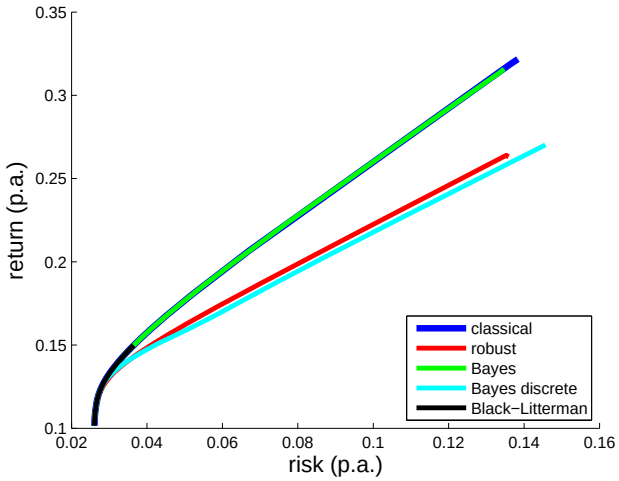


Figure 7.16: All investigated efficient frontiers on 01.11.2003.

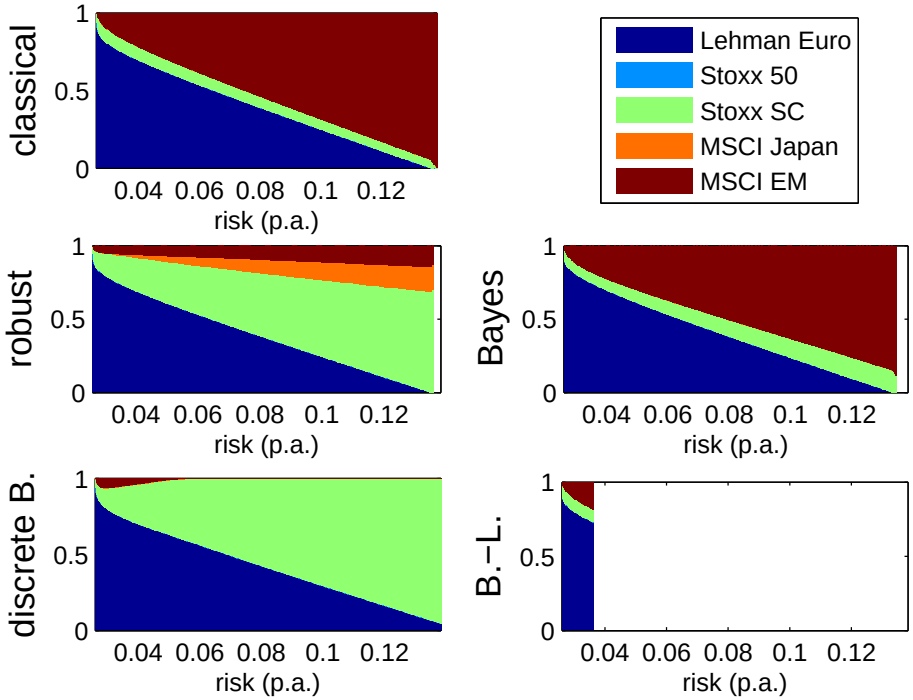


Figure 7.17: All investigated efficient portfolios on 01.11.2003.

7.3 Comparison of point estimates – Bayes vs. Black-Litterman

Both the Bayesian and the Black-Litterman approach are models that combine information from experts and from a historical data set:

- In the Bayes model the expert describes the prior distribution of the parameters which then gets conditioned on the data.
  In short: data = condition, experts = prior.

- The Black-Litterman model basically works vice versa: The prior is determined by the data sample, and afterwards different experts' opinions alter the final outcome.
  In short: data = prior, experts = condition.

First of all, it is worth noting that none of the models contains the other one as a special case. The Bayesian setting allows flexibility in modeling the covariance matrix, i.e. the covariance is supposed to follow a certain distribution, whereas in the Black-Litterman model the covariance matrix is fixed, thus no uncertainty is assumed. On the other hand, the Black-Litterman model allows more possibilities for incorporating experts' opinions. Both absolute and relative forecasts can be made on arbitrary assets. In the Bayes model however, external knowledge can only enter the model through the prior assumption which basically represents the case of making an absolute return forecast for each individual asset.

Both models can – as illustrated above – be applied in two different ways:

- Calculate point estimates for μ and Σ and perform a classical portfolio optimization using these estimates.

- Calculate the (marginal) distribution of the uncertain parameter(s), create the according uncertainty set and perform a robust portfolio optimization.

In this section we compare the point estimates for the classical optimization problem that are obtained from the two models, and in Section 7.4 we try to match the uncertainty sets for the robust portfolio optimization.

Besides comparing the general formulas given in the Bayes and the Black-Litterman model, we consider explicitly the case where the market framework in both models is the same and we hence obtain the same class of distribution for the posterior distribution of μ. In the general cases as described above, we have a Student-t distribution in the Bayes setting (see Equation (7.4)) and a normal distribution in Black-Litterman (see Equation (7.10)). The Student-t distribution stems from the prior in the Bayes model where variability in the covariance estimator is allowed – as opposed to the Black-Litterman model where the covariance matrix is assumed to be fixed. Thus, for obtaining the same class

of distriubution for the (marginal) posterior distributions of μ, we make the assumption

$$\nu_0 := \infty \tag{7.17}$$

in the Bayes model which basically reduces the prior assumption for the covariance matrix to a point estimate with zero variance. The consequences of this assumption are summarized in the following lemma.

Lemma 7.17. *Let $\nu_0 := \infty$ in the Bayesian setting. Then, the prior assumption can be rewritten as*

$$\begin{aligned} \mu &\sim \mathcal{N}\left(\mu_0, \frac{1}{d_0}\Sigma_0\right), \\ \Sigma &= \Sigma_0, \end{aligned}$$

and the marginal posterior distribution of $\mu \mid x_s, \ldots, x_S$ is given by

$$\mu \mid x_1, \ldots, x_S \sim \mathcal{N}\left(\frac{d_0}{d_0+S}\mu_0 + \frac{S}{d_0+S}\hat{\mu}, \frac{1}{d_0+S}\Sigma_0\right).$$

Proof. Recall the general formulation of the Bayes prior assumption:

$$(\mu, \Sigma) \sim \mathcal{NIW}(\mu_0, d_0, \Sigma_0, \nu_0)$$

i.e.

$$\begin{aligned} \mu \mid \Sigma &\sim \mathcal{N}\left(\mu_0, \frac{1}{d_0}\Sigma\right) \\ \Sigma &\sim \mathcal{IW}(\nu_0\Sigma_0, \nu_0 + n + 1). \end{aligned}$$

With $\nu_0 = \infty$, the last expression implies

$$\begin{aligned} \mathbf{E}[\Sigma] &= \Sigma_0, \\ \mathbf{Cov}[\Sigma] &= 0, \end{aligned}$$

i.e. this simply represents $\Sigma = \Sigma_0$. Thus, the prior assumption with $\nu_0 = \infty$ can equivalently be described by

$$\begin{aligned} \mu &\sim \mathcal{N}\left(\mu_0, \frac{1}{d_0}\Sigma_0\right), \\ \Sigma &= \Sigma_0. \end{aligned}$$

Accordingly, as a Student-t distribution with infinitely many degrees of freedom is equal to a normal distribution, the marginal posterior of $\mu \mid x_1, \ldots, x_S$ simplifies

as follows:

$$\begin{aligned}\mu \mid x_1, \ldots, x_S &\sim St\left(\mu_1, \frac{1}{d_1}\Sigma_1, \infty\right) \\ &= \mathcal{N}\left(\mu_1, \frac{1}{d_1}\Sigma_1\right) \\ &= \mathcal{N}\left(\frac{d_0}{d_0+S}\mu_0 + \frac{S}{d_0+S}\hat{\mu}, \frac{1}{d_0+S}\Sigma_0\right)\end{aligned}$$

Note that in the limit $\nu_0 = \infty$, it holds that $\Sigma_1 = \Sigma_0$. □

We want to compare the Bayes and the Black-Litterman approach with respect to both point estimates needed in the classical portfolio optimization and an uncertainty set for the return which is used in the robust portfolio optimization problem. As in both the classical and the robust portfolio optimization problem, no uncertainty of the covariance matrix is explicitly accounted for, it is first of all necessary to match the point estimates for Σ. Hence, as the Black-Litterman model simply uses the maximum likelihood estimator $\hat{\Sigma}$ as point estimate for Σ, we have to limit the choices in the Bayes model further by defining

$$\Sigma_0 := \hat{\Sigma}.$$

Notation 7.18. *As notational convention, we will call the Bayes framework with the two definitions*

- $\nu_0 := \infty$ *and*
- $\Sigma_0 := \hat{\Sigma}$

the restricted Bayes model.

7.3.1 Restricted Bayes vs. Black-Litterman

We first compare the point estimates for the vector of expected returns and the covariance matrix obtained from the restricted Bayes model and the Black-Litterman approach. Recalling the results from above, we have the following formulas for the point estimates:

$$\begin{aligned}\hat{\mu}_B &= \mathbf{E}[\mu \mid x_1, \ldots, x_S] \\ &= \frac{d_0}{d_0+S}\mu_0 + \frac{S}{d_0+S}\hat{\mu}, \\ \hat{\Sigma}_B &= \hat{\Sigma},\end{aligned}$$

and

$$\begin{aligned}\hat{\mu}_{BL} &= \mathbf{E}[\mu \mid Q = q] \\ &= \hat{\mu} + \tau\hat{\Sigma}P^T(\tau \cdot P\hat{\Sigma}P^T + \Omega)^{-1}(q - P\hat{\mu}) \\ &= \hat{\mu} + \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}(q - P\hat{\mu}), \\ \hat{\Sigma}_{BL} &= \hat{\Sigma}.\end{aligned}$$

As it holds that $\hat{\Sigma}_B = \hat{\Sigma}_{BL} = \hat{\Sigma}$, it suffices to compare the point estimates for the return vector. For testing coincidence of $\hat{\mu}_B$ and $\hat{\mu}_{BL}$, we analyze if a given Bayes prior can be expressed as a Black-Litterman forecast such that the resulting estimates are the same, and vice versa. In the Bayes model, the free variables are μ_0 and d_0, and in the Black-Litterman model, we can choose P, q and τ. Recall that the matrix Ω was assumed to be given by $\Omega = (1-\tau)P\hat{\Sigma}P^T$.

We will find that when simply considering point estimates for the classical portfolio optimization, each model can be expressed as a special case of the other one by adequately defining the free parameters. These results are summarized in the following two propositions.

Proposition 7.19. *Let the restricted Bayes model be given, i.e. μ_0 and d_0 are fixed. Then, there exist parameters P, q and τ such that the point estimates for the classical portfolio optimization problem coincide in the restricted Bayes and the Black-Litterman model. Hence, the restricted Bayes model is a special case of the Black-Litterman model.*

Proof. To show that an arbitrary choice of the priors μ_0 and d_0 in the restricted Bayes model is contained as a special case in the Black-Litterman model, we equate the formulas for the (posterior) point estimates and determine working values for the free variables P, q and τ:

$$\hat{\mu} + \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}(q - P\hat{\mu}) \overset{!}{=} \frac{d_0}{d_0+S}\mu_0 + \frac{S}{d_0+S}\hat{\mu}. \tag{7.18}$$

Since this is one equation for three free variables, appropriate definitions for P, q and τ are possible. Restricting the choices for P to invertible matrices, further equivalent reformulations can be performed:

$$\begin{aligned}\Leftrightarrow \quad & \hat{\mu} + \tau P^{-1}(q - P\hat{\mu}) = \frac{d_0}{d_0+S}\mu_0 + \frac{S}{d_0+S}\hat{\mu} \\ \Leftrightarrow \quad & (1-\tau)\hat{\mu} + \tau P^{-1}q = \frac{d_0}{d_0+S}\mu_0 + \frac{S}{d_0+S}\hat{\mu} \\ \Leftrightarrow \quad & q = \frac{1}{\tau}P\left(\frac{d_0}{d_0+S}\mu_0 + \left[\frac{S}{d_0+S} - (1-\tau)\right]\hat{\mu}\right) \\ & = P\left[\left(1 - \tau\frac{d_0}{d_0+S}\right)\hat{\mu} + \tau\frac{d_0}{d_0+S}\mu_0\right]. \end{aligned} \tag{7.19}$$

Thus, any definitions for an invertible matrix P, a vector q and a positive real number τ fulfilling the above formula (7.19) can be used to express a given prior assumption for μ_0 and d_0 in the Black-Litterman framework and obtain the same point estimates for the classical portfolio optimization problem.

A canonical choice for P, q and τ satisfying the relationship in Equation (7.19) is

$$P := I, \quad q := \mu_0, \quad \tau := \frac{d_0}{d_0 + S}.$$

These parameter choices seem rather intuitive: As in the Bayes prior a particular vector μ_0 for the returns and nothing else is given, it is natural to express this vector as an absolute forecast for each asset in the Black-Litterman model, i.e. $P = I$ and $q = \mu_0$. The scaling factor τ is finally adjusted such that the formulas are equal. □

Proposition 7.20. *Let the Black-Litterman model be given, i.e. P, q and τ are fixed. Then, the prior parameters μ_0 and d_0 in the restricted Bayes model can be defined such that the point estimates in the given Black-Litterman and the restricted Bayes model coincide. Thus, the Black-Litterman model is a special case of the restricted Bayes model with respect to comparison of the point estimates.*

Proof. We again compare the formulas for the (posterior) point estimates in both models, see Equation (7.18) in the proof of the previous proposition. Letting $d_0 > 0$ be chosen arbitrarily and solving this equation for the prior vector μ_0, we obtain

$$\begin{aligned}\mu_0 &= \frac{d_0 + S}{d_0} \cdot \left[\hat{\mu} - \frac{S}{d_0 + S}\hat{\mu} + \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}(q - P\hat{\mu})\right] \\ &= \hat{\mu} + \frac{d_0 + S}{d_0}\tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}(q - P\hat{\mu}).\end{aligned}$$

Thus, defining the prior μ_0 in such a way, i.e. any forecasts (absolute or relative) are contained therein, we have incorporated arbitrary experts' opinions into the existing restricted Bayesian framework and achieve the same point estimate for the posterior return as we would get in the Black-Litterman model itself. This approach hence uses the sound statistical method of Bayes while not requiring to come up with a particular point estimate for the entire return vector, as maybe we do not have an opinion about some of the assets. □

To summarize, when comparing the restricted Bayesian setting (i.e. without variability in the covariance matrix) and the Black-Litterman approach with respect to the final point estimates for the classical portfolio optimization problem, we have seen that each model can be expressed as a special case of the other one.

7.3.2 (General) Bayes vs. Black-Litterman

We now want to compare the point estimates determined by the general Bayesian model (i.e. $\nu_0 \neq \infty$) and the Black-Litterman approach. From the previous sections it is known that the posterior distributions of μ are different in the two settings: in Bayes a Student-t distribution is obtained, and in Black-Litterman $\mu \mid Q = q$ is normally distributed. Hence, the models cannot completely coincide (in case of a finite number of degrees of freedom in the Student-t distribution), but we can still test whether it is possible that the resulting point estimates are the same.

Recalling the respective point estimates in both the Bayes (see Equations (7.5) and (7.6)) and the Black-Litterman (Equations (7.12) and (7.13)) model, we have

$$\begin{aligned}
\hat{\mu}_B &= \mathbf{E}[\mu \mid x_1, \ldots, x_S] \\
&= \frac{d_0}{d_0 + S}\mu_0 + \frac{S}{d_0 + S}\hat{\mu}, \\
\hat{\Sigma}_B &= \mathbf{E}[\Sigma \mid x_1, \ldots, x_S] \\
&= \frac{1}{\nu_0 + S - n - 1}\left[S\hat{\Sigma} + \nu_0\Sigma_0 + \frac{d_0 S}{d_0 + S}(\mu_0 - \hat{\mu})(\mu_0 - \hat{\mu})^T\right]
\end{aligned}$$

and

$$\begin{aligned}
\hat{\mu}_{BL} &= \mathbf{E}[\mu | Q = q] \\
&= \hat{\mu} + \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}(q - P\hat{\mu}), \\
\hat{\Sigma}_{BL} &= \hat{\Sigma}.
\end{aligned}$$

Since in the Black-Litterman model the covariance estimate is given by $\hat{\Sigma}$ and the Bayes model allows modification in the covariance, it can be stated that the Bayes model cannot be contained in the Black-Litterman model as a special case. Note that this would only be true in the particular case of $\hat{\Sigma}_B = c \cdot \hat{\Sigma}$, i.e. if the Bayesian estimate for the covariance matrix a multiple of the matrix $\hat{\Sigma}$ which is determined by the data sample. We will not analyze this case explicitly, as it basically reduces to the restricted Bayesian framework.

Vice versa, the following proposition shows that the Black-Litterman model can be regarded as a special case of the (general) Bayes model by choosing the prior parameters μ_0 and Σ_0 appropriately.

Proposition 7.21. *Let the Black-Litterman model be given, i.e. P, q and τ are fixed. Then, the prior parameters μ_0, d_0, ν_0 and Σ_0 in the (general) Bayes model can be defined such that the point estimates in the Black-Litterman and the Bayes model coincide, i.e.*

$$\hat{\mu}_B = \hat{\mu}_{BL} \quad \text{and} \quad \hat{\Sigma}_B = \hat{\Sigma}_{BL} = \hat{\Sigma}.$$

Hence, the Black-Litterman model is a special case of the Bayes model with respect to comparison of the point estimates.

Proof. Note that since the restricted Bayes model is a special case of the general Bayes model, the choice of $\nu = \infty$, $\Sigma_0 = \hat{\Sigma}$ and μ_0 as in Proposition 7.20 trivially gives the desired match of the point estimates. But also in case of $\nu \neq \infty$, appropriate parameter definitions are possible.

First, we set the equations for the return point estimates equal and solve for the variable μ_0 while $d_0 > 0$ is arbitrary. This was already done in Proposition 7.20 and yields

$$\mu_0 = \hat{\mu} + \frac{d_0 + S}{d_0} \tau \hat{\Sigma} P^T (P \hat{\Sigma} P^T)^{-1} (q - P\hat{\mu}).$$

Second, letting $\nu_0 > 0$ be arbitrarily chosen and equating the formulas for the covariance estimates gives, when solving for Σ_0:

$$\Sigma_0 = \frac{\nu_0 - n - 1}{\nu_0} \hat{\Sigma} - \frac{d_0 S}{\nu_0 (d_0 + S)} (\mu_0 - \hat{\mu})(\mu_0 - \hat{\mu})^T.$$

Hence, defining the prior variables μ_0 and Σ_0 as stated, we achieve equality of the point estimates $\hat{\mu}_B = \hat{\mu}_{BL}$ and $\hat{\Sigma}_B = \hat{\Sigma}_{BL} = \hat{\Sigma}$. □

Thus, similar to the restricted Bayes model, we can use the statistical method of the (general) Bayes framework and include relative forecasts by defining the prior assumptions accordingly.

Altogether, we have found in this section that the general Bayes model contains the Black-Litterman approach as a special case if the focus is on the classical portfolio optimization where point estimates for the parameters μ and Σ are needed. As the Bayes model allows variability in the structure of the covariance matrix and in the Black-Litterman model the covariance estimate is given by $\hat{\Sigma}$, it is hence not possible to define forecasts in the Black-Litterman approach such that the Bayes model merely represents a special case.

7.4 Comparison of uncertainty sets – Bayes vs. Black-Litterman

In this section we analyze if the parameters needed for the robust portfolio optimization can be matched in the models of Bayes and Black-Litterman. Hence, it is necessary to compare the uncertainty set for the return vector (i.e. the midpoint, the shape matrix and the size) and the point estimate for the covariance matrix.

For ease of comparison, we summarize all the needed formulas from the previous sections where $\hat{\mu}$ and $\hat{\Sigma}$ with the respective subscripts denote the point

estimates and m and C the midpoint and the shape matrix defining the uncertainty sets.

$$\begin{aligned}
\hat{\mu}_B &= \mathbf{E}[\mu \mid x_1, \ldots, x_S] = \frac{d_0}{d_0 + S}\mu_0 + \frac{S}{d_0 + S}\hat{\mu}, \\
\hat{\Sigma}_B &= \mathbf{E}[\Sigma \mid x_1, \ldots, x_S] \\
&= \frac{1}{\nu_0 + S - n - 1}\left[S\hat{\Sigma} + \nu_0\Sigma_0 + \frac{d_0 S}{d_0 + S}(\mu_0 - \hat{\mu})(\mu_0 - \hat{\mu})^T\right], \\
m_B &= \mathbf{E}[\mu \mid x_1, \ldots, x_S] = \hat{\mu}_B = \frac{d_0}{d_0 + S}\mu_0 + \frac{S}{d_0 + S}\hat{\mu}, \\
C_B &= \mathbf{Cov}[\mu \mid x_1, \ldots, x_S] \\
&= \frac{1}{\nu_0 + S - n - 1} \cdot \frac{1}{d_0 + S}\left[S\hat{\Sigma} + \nu_0\Sigma_0 + \frac{d_0 S}{d_0 + S}(\mu_0 - \hat{\mu})(\mu_0 - \hat{\mu})^T\right]
\end{aligned}$$

and

$$\begin{aligned}
\hat{\mu}_{BL} &= \mathbf{E}[\mu | Q = q] = \hat{\mu} + \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}(q - P\hat{\mu}), \\
\hat{\Sigma}_{BL} &= \hat{\Sigma}, \\
m_{BL} &= \mathbf{E}[\mu | Q = q] = \hat{\mu}_{BL} = \hat{\mu} + \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}(q - P\hat{\mu}), \\
C_{BL} &= \mathbf{Cov}[\mu | Q = q] = \tau\hat{\Sigma} - \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}\tau P\hat{\Sigma}.
\end{aligned}$$

Recall as well that in the restricted Bayes model, the estimators $\hat{\Sigma}_B$ and C_B reduce to

$$\hat{\Sigma}_B = \hat{\Sigma} \quad \text{and} \quad C_B = \frac{1}{d_0 + S}\hat{\Sigma}.$$

In mathematical terms, in Section 7.3 we compared and matched the expressions

$$\hat{\mu}_B \stackrel{!}{=} \hat{\mu}_{BL} \quad \text{and} \quad \hat{\Sigma}_B \stackrel{!}{=} \hat{\Sigma}_{BL}$$

while in this section we want to match

$$\mathcal{U}_B \stackrel{!}{=} \mathcal{U}_{BL} \quad \text{and} \quad \hat{\Sigma}_B \stackrel{!}{=} \hat{\Sigma}_{BL}$$

which can equivalently be replaced by

$$m_B \stackrel{!}{=} m_{BL}, \quad C_B \stackrel{!}{=} C_{BL} \quad \text{and} \quad \hat{\Sigma}_B \stackrel{!}{=} \hat{\Sigma}_{BL}$$

since for elliptical distributions equal moments will result in the same uncertainty set (with possibly different sizes δ though) when defined by confidence ellipsoids.

Analogous to the previous section, we first compare the easier setting of the restricted Bayes model to the Black-Litterman framework before extending the investigation to the general Bayesian approach.

7.4.1 Restricted Bayes vs. Black-Litterman

Since in the restricted Bayes model the point estimate for Σ is given by $\hat{\Sigma}$, i.e. it coincides with the estimate in the Black-Litterman model, it suffices in this case to compare the uncertainty set for the return. Furthermore, both posterior distributions are given in terms of a normal distribution which implies that the sizes of the respective uncertainty sets are determined as quantiles of a χ^2_n-distribution to the appropriate confidence level. Hence, the uncertainty sets have the same size and it remains to equate the formulas for the midpoint and the shape matrix.

Analogous to the investigations in the previous section, we distinguish the two cases of trying to express the Bayes prior as Black-Litterman forecast and vice versa.

Proposition 7.22. *Let the restricted Bayes model be given, i.e. μ_0 and d_0 are fixed. Then there exist parameters P, q and τ such that the uncertainty sets obtained from the restricted Bayes model and the Black-Litterman approach are the same. Hence, with respect to matching parameters for the robust portfolio optimization problem, the restricted Bayes model is a special case of Black-Litterman.*

Proof. Given the prior assumptions μ_0 and d_0, we want to determine P, q and τ such that

$$\hat{\mu} + \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}(q - P\hat{\mu}) \stackrel{!}{=} \frac{d_0}{d_0+S}\mu_0 + \frac{S}{d_0+S}\hat{\mu}$$

and

$$\tau\hat{\Sigma} - \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}\tau P\hat{\Sigma} \stackrel{!}{=} \frac{1}{d_0+S}\hat{\Sigma}$$

hold. Choosing again P as an invertible matrix, we already know from the calculations in the proof of Proposition 7.19 that the first equation leads to

$$q = P\left[\left(1 - \tau\frac{d_0}{d_0+S}\right)\hat{\mu} + \tau\frac{d_0}{d_0+S}\mu_0\right].$$

The second equation simplifies to

$$\tau\hat{\Sigma} - \tau^2\hat{\Sigma} \stackrel{!}{=} \frac{1}{d_0+S}\hat{\Sigma}$$

which leads to the two solutions for τ:

$$\tau = \frac{1 \pm \sqrt{1 - 4\frac{1}{d_0+S}}}{2}.$$

Note that both possible values are in $[0,1]$ since $\sqrt{1 - 4\frac{1}{d_0+S}} \leq 1$. With τ known, the appropriate value for q can be determined by the above equation, hence

it holds that any Bayes prior μ_0, d_0 can be represented as a Black-Litterman forecast such that the resulting posterior distributions of the return vector are the same. □

Proposition 7.23. *Let the Black-Litterman setting be given, i.e. the parameters P, q and τ and assume that P is invertible. Then, the parameters μ_0 and d_0 in the restricted Bayes model can be defined appropriately such that the uncertainty sets coincide, i.e. the Black-Litterman approach is a special case of the restricted Bayes model with respect to comparison of the uncertainty sets.*

Proof. In case of P invertible, the formulas for the Black-Litterman estimates simplify to

$$\begin{aligned} m_{BL} &= \hat{\mu} + \tau P^{-1}(q - P\hat{\mu}) = (1-\tau)\hat{\mu} + \tau P^{-1} q, \\ C_{BL} &= \tau(1-\tau)\hat{\Sigma}. \end{aligned}$$

Hence, from setting $m_{BL} = m_B$ we obtain

$$\begin{aligned} \mu_0 &= \frac{d_0 + S}{d_0}\left[\left(1 - \tau - \frac{S}{d_0 + S}\right)\hat{\mu} + \tau P^{-1} q\right] \\ &= \frac{d_0 + S}{d_0}\left[\left(\frac{d_0}{d_0 + S} - \tau\right)\hat{\mu} + \tau P^{-1} q\right] \\ &= \left(1 - \tau\,\frac{d_0 + S}{d_0}\right)\hat{\mu} + \tau\,\frac{d_0 + S}{d_0} P^{-1} q. \end{aligned}$$

Equating additionally $C_{BL} = C_B$ finally yields

$$d_0 = \frac{1}{\tau(1-\tau)} - S.$$

Therefore, the Black-Litterman model with an invertible matrix P can be expressed as a special case of the restricted Bayes model. □

7.4.2 (General) Bayes vs. Black-Litterman

To compare the general Bayes model and the Black-Litterman approach with respect to the parameters needed in the robust portfolio optimization, the uncertainty set for the return vector and the point estimate for the covariance matrix have to be investigated. In the Bayes model the posterior of μ is given by a Student-t distribution whereas in the Black-Litterman setting the posterior is described by a normal distribution. As both the normal and the Student-t distribution are elliptical, the respective uncertainty sets are created using the first two moments of the posterior distribution of μ. Hence, by equating the moments, it could be achieved that the midpoint and the shape are the same, but the uncertainty sets cannot be easily matched completely, as the quantiles to determine the

sizes have to be obtained from either an F-distribution (in the Bayesian model where $\mu \mid x_1, \ldots, x_S \sim St$) or a χ^2-distribution (in the Black-Litterman model where $\mu \mid Q = q \sim \mathcal{N}$). Therefore, setting a particular confidence level results in differently sized uncertainty sets, or vice versa, fixing the same size for the two ellipsoids, they correspond to different levels of confidence.

Neglecting the size for a moment and investigating only the midpoint and the shape of the uncertainty set and the point estimate for the covariance, the following three equations have to hold:

$$m_B = m_{BL}, \quad C_B = C_{BL} \quad \text{and} \quad \hat{\Sigma}_B = \hat{\Sigma}_{BL},$$

i.e. we obtain the equations

$$\begin{aligned}
&(1)\ \frac{d_0}{d_0+S}\mu_0 + \frac{S}{d_0+S}\hat{\mu} = \hat{\mu} + \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}(q - P\hat{\mu}),\\
&(2)\ \frac{1}{\nu_0+S-n-1}\cdot\frac{1}{d_0+S}\left[S\hat{\Sigma} + \nu_0\Sigma_0 + \frac{d_0 S}{d_0+S}(\mu_0-\hat{\mu})(\mu_0-\hat{\mu})^T\right]\\
&\qquad = \tau\hat{\Sigma} - \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}\tau P\hat{\Sigma},\\
&(3)\ \frac{1}{\nu_0+S-n-1}\left[S\hat{\Sigma} + \nu_0\Sigma_0 + \frac{d_0 S}{d_0+S}(\mu_0-\hat{\mu})(\mu_0-\hat{\mu})^T\right] = \hat{\Sigma}.
\end{aligned}$$

As in Section 7.3.2 when comparing the point estimates, we straightforwardly have that the Bayes model (in general) cannot be expressed in the Black-Litterman framework since in the Bayes model variability in the covariance is allowed whereas in the Black-Litterman setting the covariance is given by $\hat{\Sigma}$, i.e. for arbitrarily chosen parameters in the Bayesian prior, Equation (3) does not hold. Vice versa, we assume that the Black-Litterman model is given. From Equation (1) the prior parameters μ_0 can be determined. Defining furthermore Σ_0 such that Equation (3) is satisfied (by simply solving for Σ_0, see Proposition 7.21), Equation (3) can be plugged into (2) and thus yields

$$\frac{1}{d_0+S}\hat{\Sigma} = \tau\hat{\Sigma} - \tau\hat{\Sigma}P^T(P\hat{\Sigma}P^T)^{-1}\tau P\hat{\Sigma}.$$

Assuming additionally P to be an invertible matrix, the equation simplifies to the restriction

$$\frac{1}{d_0+S}\hat{\Sigma} = \tau(1-\tau)\hat{\Sigma}$$

which gives $d_0 = \frac{1}{\tau(1-\tau)} - S$.

These conclusions are summarized in the following proposition.

Proposition 7.24. *Let the Black-Litterman model be given and assume P to be invertible. Then, the prior parameters μ_0, d_0, ν_0 and Σ_0 in the (general) Bayes model can be defined such that the midpoint and the shape of the uncertainty set and the point estimate for the covariance coincide in the Black-Litterman and the Bayes model.*

7.5 Summary of the comparisons

The following Table 7.2 summarizes the results in Sections 7.3 and 7.4, where the notation "$\subset$" is supposed to be read as "can be expressed as a special case of":

Comparison of point estimates		
restricted Bayes	$\subset$	Black-Litterman
Black-Litterman	$\subset$	restricted Bayes
Bayes	$\not\subset$	Black-Litterman
Black-Litterman	$\subset$	Bayes
Comparison of uncertainty sets		
restricted Bayes	$\subset$	Black-Litterman
Black-Litterman with P invertible	$\subset$	restricted Bayes
Bayes	$\not\subset$	Black-Litterman
Black-Litterman with P invertible	$\subset$	Bayes

Table 7.2: Summary of the relationships between the Bayes and the Black-Litterman model.

Chapter 8

Summary and Outlook

This dissertation is organized in two parts. The first part consists of Chapters 2 and 3 and investigates theoretical aspects of a general convex conic optimization problem and its associated local robust counterpart. As the main point of interest was uniqueness and stability (i.e. continuity) of the optimal solution, the results for the original problem were summarized and proved before analogous analyses were performed for the robust problem. To the best of our knowledge, these investigations for the robust counterpart have not been done so far. We found that the same stability properties hold for the robust problem as for the original one. Furthermore, we could show that in a rather general formulation of the original problem where the uncertain parameter enters the objective function linearly, an ellipsoidal uncertainty set is more promising as a polyhedral one, as it results in a certain structure of the set of optimal solutions. If additionally a constraint is imposed that does not allow multiples of the optimal solution, uniqueness of $x^*(u)$ is guaranteed, and hence continuity follows. With respect to the costs of robustification, we were able to prove the result of Ben-Tal and Nemirovski [4] for linear programs in our general conic setting, namely that the increase in the optimal objective value depends linearly on the size of the respective uncertainty set.

The second part of the dissertation contains Chapters 4 to 7 and illustrates the application and the benefits of the robust counterpart to the well-known portfolio optimization problem of Markowitz. As the uncertain parameters in this problem are the vector of expected assets returns and the covariance matrix thereof which are usually estimated from a sample of historical observations and thus contain estimation errors, the application of the robust counterpart approach seems promising.

To solve the associated robust program, we discussed two different ideas to describe appropriate uncertainty sets. We first defined an uncertainty set using a confidence ellipsoid which lead to the rather surprising result that the robust efficient frontier is identical to the classical one, but shortened. Even though confidence ellipsoids are rather natural choices for uncertainty sets, this result

seems to have been unnoticed so far. Building an uncertainty set based on various statistical estimators – which could all serve as point estimates equally likely – for the mean of an elliptical distribution gives robust portfolios which illustrate the effects of robustification quite nicely.

We furthermore investigated consistency of the input parameters and especially of the resulting optimal portfolios, i.e. we studied the case of having an infinitely large sample of historical data. There, we could prove by using uniqueness and hence continuity of the optimal solutions that both the classical and the robust optimal portfolio are consistent estimators for the true (but unknown) portfolio in case of consistent parameter estimates.

Finally, more sophisticated methods of obtaining suitable point estimates and uncertainty sets were considered in the last chapter. The Bayesian and the Black-Litterman approach are models including both information of an existing sample and external input, either given in form of an assumed prior distribution or in terms of explicit individual forecasts.

Summarizing, using the Bayesian or the Black-Litterman approach seems only sensible if there is a strong belief in the prior assumption. Otherwise, the approach of using several statistical estimators seems rather promising as it exploits the available data sample and creates more diversified portfolio allocations.

Besides the models of Bayes and Black-Litterman, there is a further approach to include prior knowledge into the parameter determination which was introduced by Qian and Gorman [66]. Their model basically extends the Black-Litterman approach to additionally allowing variability in and forecasts for the coavariance matrix. It starts with modelling the market itself according to a normal distribution, i.e. $X \sim \mathcal{N}(\hat{\mu}, \hat{\Sigma})$, and then describes forecasts analogously to the Black-Litterman approach using a projection matrix P. It is worth noting that in the Qian-Gorman model the prior assumption is based on the distribution of the actual asset returns X, whereas both the Bayes and the Black-Litterman model are working on the distribution of the *expected* returns μ. This fact leads to a major drawback of the Qian-Gorman approach when using it for determining parameters for the optimization problems. Due to the framework being based on the market returns X, only *point estimates* for the parameters μ and Σ can be obtained from the posterior distribution of X. The posterior *distribution* of μ (or the first two moments thereof) is not available, but would be needed to create an uncertainty set. Hence, the Qian-Gorman approach can only be used if the classical portfolio optimization problem is solved, but it does not yield an uncertainty set which could be applied in the robust optimization.

A further aspect that could be worth investigating more closely is the size of the uncertainty set. It is known that and how it influences the costs of robustification. But the size also describes the level of conservativeness imposed on the problem: Setting δ large, the optimal solution might be too conservative and the costs will be unnecessarily high. On the other hand, δ small can first of all lead to numerical difficulties and thus falsify the results, and secondly the robust

counterpart approach might not be sufficiently effective in robustifying the solution. Hence it might be worth analyzing if there is something like an "optimal size δ" describing the best trade-off between robustness and costs. This could be generally expressed by

$$\delta_{opt} = \arg\min_{\delta} \quad \mathbf{E}\left[\|x^*_{rob}(\hat{\mu}, \hat{\Sigma}, \delta) - x^*_{cl}(\mu, \Sigma)\|^2\right]$$

where $x^*_{rob}(\hat{\mu}, \hat{\Sigma}, \delta)$ denotes the optimal solution to the robust program with an uncertainty set specified by the parameters $\hat{\mu}$ and $\hat{\Sigma}$ and the size δ, and $x^*_{cl}(\mu, \Sigma)$ describes the optimal solution obtained when using the original (unknown) parameters μ and Σ. Introducing the expression $\mathbf{E}[x^*_{rob}(\hat{\mu}, \hat{\Sigma}, \delta)]$, we can reformulate[1] the above to

$$\begin{aligned}
\delta_{opt} &= \arg\min_{\delta} \quad \mathbf{E}\left[\|x^*_{rob}(\hat{\mu}, \hat{\Sigma}, \delta) - x^*_{cl}(\mu, \Sigma)\|^2\right] \\
&= \arg\min_{\delta} \quad \underbrace{\mathbf{E}\left[\|x^*_{rob}(\hat{\mu}, \hat{\Sigma}, \delta) - \mathbf{E}[x^*_{rob}(\hat{\mu}, \hat{\Sigma}, \delta)]\|^2\right]}_{\text{estimation variance}} \\
&\qquad + \underbrace{\|\,\mathbf{E}[x^*_{rob}(\hat{\mu}, \hat{\Sigma}, \delta)] - x^*_{cl}(\mu, \Sigma)\|^2}_{\text{bias}}
\end{aligned}$$

The first term, called estimation variance, expresses how much the robust solutions deviate from their expected value. The larger δ, the smaller the estimation variance will be. This holds since when choosing a large uncertainty set using the first parameter estimate, the intersection of it with the uncertainty set using a second parameter estimate will be quite large, and thus the corresponding optimal solutions of the robust counterpart program will not be very different – if they are not even identical.

The second term, called bias, expresses how much the expected robust solution differs from the classical solution calculated using the real parameters. It holds that with increasing δ, the bias also increases. This can be explained by the fact that the larger we choose δ, the more conservative the solution will be since more

[1] For readability and ease of notation, let $x^*_{rob} := x^*_{rob}(\hat{\mu}, \hat{\Sigma}, \delta)$ and $x^*_{cl} = x^*_{cl}(\mu, \Sigma)$. Note that the random variables are $\hat{\mu}$ and $\hat{\Sigma}$, hence, with respect to the expectation, x^*_{cl} is a constant. Then it holds that

$$\begin{aligned}
&\mathbf{E}\left[\|x^*_{rob} - x^*_{cl}\|^2\right] \\
&\quad = \mathbf{E}\left[\left(x^*_{rob} - x^*_{cl}\right)^T \left(x^*_{rob} - x^*_{cl}\right)\right] \\
&\quad = \mathbf{E}\left[\left(x^*_{rob} - \mathbf{E}[x^*_{rob}] + \mathbf{E}[x^*_{rob}] - x^*_{cl}\right)^T \cdot \left(x^*_{rob} - \mathbf{E}[x^*_{rob}] + \mathbf{E}[x^*_{rob}] - x^*_{cl}\right)\right] \\
&\quad = \mathbf{E}\left[\|x^*_{rob} - \mathbf{E}[x^*_{rob}]\|^2\right] + \|\,\mathbf{E}[x^*_{rob}] - x^*_{cl}\|^2 + 2\underbrace{\mathbf{E}\left[x^*_{rob} - \mathbf{E}[x^*_{rob}]\right]^T}_{=0} \cdot \left(\mathbf{E}[x^*_{rob}] - x^*_{cl}\right).
\end{aligned}$$

possible parameter realizations have to be taken into account. Thus, the robust solution will differ more from $x^*_{cl}(\mu, \Sigma)$ for increasing size of the uncertainty set.

These opposing goals could be used to determine an optimal size δ, but at the moment it is not yet clear if the optimal δ always is a strictly positive number, or if the infimum is zero. A further problem is that the real market parameters are still unknown, but enter the optimization problem to find δ. An approximation could be obtained by simulations, but it would have to be analyzed if this falsifies the results. Still, this seems to be an approach worth pursuing.

Appendix A

Convex analysis

In the first part of this appendix we summarize definitions of local and global Lipschitz continuity and give a useful result linking some of them. Afterwards, we give the definitions of different types of directional differentiability and some useful properties thereof.

Definition A.1 (pointwise Lipschitz continuity). *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called* pointwise Lipschitz continuous *in x if and only if there exists a neighborhood $V(x)$ around x and a constant $L = L(x) > 0$ such that*

$$\|f(x) - f(y)\| \leq L \cdot \|x - y\| \quad \forall y \in V(x).$$

Definition A.2 (local Lipschitz continuity in a point). *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is* locally Lipschitz continuous *in x if and only if there exists a neighborhood $V(x)$ around x and a constant $L = L(x) > 0$ such that*

$$\|f(a) - f(b)\| \leq L \cdot \|a - b\| \quad \forall a, b \in V(x).$$

Obviously we have that local Lipschitz continuity in x implies pointwise Lipschitz continuity in x. The other direction does not holds as shown by the counterexample $f(x) = x \ \sin\left(\frac{1}{x}\right)$, which is pointwise Lipschitz continuous for all $x \in \mathbb{R}$ but not locally Lipschitz continuous in $x = 0$.

Definition A.3 (local Lipschitz continuity). *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is* locally Lipschitz continuous *if and only if f is locally Lipschitz continuous in x for all $x \in \mathbb{R}^n$.*

Definition A.4 (global Lipschitz continuity). *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is* globally Lipschitz continuous *if and only if there exists a constant $L > 0$ such that*

$$\|f(x) - f(y)\| \leq L \cdot \|x - y\| \quad \forall x, y \in \mathbb{R}^n.$$

Note that in the definition of global Lipschitz continuity the constant L must be the same for all x and thus independent of any neighborhoods.

Lemma A.5. *If a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then it is also locally Lipschitz continuous.*

Proof. For a proof see Rockafellar [70], p. 86, Theorem 10.4. □

Obviously the same statement holds for concave functions.

At some places in this dissertation we need the notion of directional differentiability and results based thereupon. The definitions and statements here can be found e.g. in the book of Bonnans and Shapiro [18], Section 2.2.1.

We assume $f : \mathbb{R}^n \to \mathbb{R}^m$ to be an arbitrary mapping.

Definition A.6. *The function f is said to be directionally differentiable at a point $x \in \mathbb{R}^n$ in a direction $d \in \mathbb{R}^n$ if*

$$f'(x;d) := \lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t} < \infty.$$

If f is directionally differentiable at x in every direction $d \in \mathbb{R}^n$, f is said to be directionally differentiable at x.

Proposition A.7. *If the directional derivative $f'(x;d)$ exists, it is positively homogeneous in d, i.e.*

$$f'(x;\alpha d) = \alpha f'(x;d) \quad \forall \alpha \geq 0.$$

Proof. See Bonnans and Shapiro [18], page 34. □

The definition for Hadamard directional differentiability is an even stronger concept:

Definition A.8. *The function f is said to be Hadamard directionally differentiable at a point $x \in \mathbb{R}^n$ if it is directionally differentiable in x (i.e. the directional derivative $f'(x;d)$ exists for all directions $d \in \mathbb{R}^n$) and it holds that*

$$f'(x;d) = \lim_{\substack{t \downarrow 0 \\ \tilde{d} \to d}} \frac{f(x+t\tilde{d}) - f(x)}{t}.$$

If additionally $f'(x;d)$ is linear in d, f is said to be Hadamard differentiable at x.

Note that the special case of $\tilde{d} = d$ reduces the Hadamard directional differentiability to directional differentiability as given in Definition A.6.

Proposition A.9. *If f is Hadamard directionally differentiable at $x \in \mathbb{R}^n$, then the directional derivative $f'(x;.)$ is continuous on $\mathbb{R}^n$.*

Proof. See Bonnans and Shapiro [18], Proposition 2.46. □

Another notion of directional differentiability is Fréchet directional differentiability:

Definition A.10. *The function f is said to be Fréchet directionally differentiable at a point $x \in \mathbb{R}^n$ if it is directionally differentiable in x and it holds that*

$$f(x+d) = f(x) + f'(x;d) + o(\|d\|)$$

for any $d \in \mathbb{R}^n$.

In finite dimensional spaces[1], i.e. in $\mathbb{R}^n$, these concepts of differentiability are closely linked to one another:

Proposition A.11. *The following statements hold:*

(i) Hadamard directional differentiability implies Fréchet directional differentiability.

(ii) Fréchet directional differentiability together with continuity of $f'(x;.)$ implies Hadamard directional differentiability.

(iii) If f is locally Lipschitz continuous, then Hadamard and Fréchet directional differentiability are equivalent.

Proof. See Bonnans and Shapiro [18], page 36. □

A particular result involving directional derivatives and which we need in the proof of Theorem 3.37 is the following:

Proposition A.12. *Consider the following type of optimization problem:*

$$\begin{aligned} \min_{x \in X} \quad & f(x) \\ s.t. \quad & G(x) + u \leq_K 0 \end{aligned} \tag{P}$$

with f being convex and G being K-convex in x. The problem with $u = 0$ (i.e. the unperturbed problem) will be denoted by (P_0), and its dual[2] with (D_0). Assume furthermore that the feasibility set $\mathcal{F}_P^*(u)$ is non-empty for all u in a neighborhood of $\hat{u} = 0$ (equivalently assume the existence of a Slater point for (P_0)), and that the optimal value $f_P^*(0)$ is finite.

[1]The definitions of directional differentiability in different senses are usually given in more general terms where f is a mapping from one (linear) normed space to another. In our setting it suffices to consider finite real spaces $\mathbb{R}^n$ and $\mathbb{R}^m$.

[2]The dual problem is given by

$$\max_{v \in K^*} \min_{x \in X} f(x) + v^T\big(G(x) + u\big),$$

but note that we do not need the explicit formulation, we only need to refer to it.

Then it holds that

(i) The set of optimal solutions $\mathcal{F}^*_{D_0} \subset K^*$ of the dual problem (D_0) is nonempty and bounded;

(ii) The optimal value function $f^*_P(u)$ is continuous at $u = 0$;

(iii) The optimal value function $f^*_P(u)$ is Hadamard directionally differentiable at $u = 0$ and it holds that

$$f^{*'}_P(0; d) = \max_{v \in \mathcal{F}^*_{D_0}} v^T d$$

for all directions d.

Proof. See Bonnans and Shapiro [16], Theorem 4.2. □

Corollary A.13. *Let the prerequisites of Proposition A.12 hold. We then especially obtain that $f^{*'}_P(0; d)$ is finite for all directions d.*

Proof. Finiteness of $f^{*'}_P(0; d)$ follows from the definition of Hadamard differentiability of $f^*_P(u)$ at $u = 0$ which in turn is given according to Proposition A.12, part (iii). □

Appendix B

Hausdorff distance

There are different possibilities to define distances. We will use the definition that the distance between a point x and a set A is the distance beween x and its projection onto A, i.e. the distance between x and the point within A which is closest to x. Evidently, the distance between any point within the set and the set itself is zero.

Definition B.1 (Hausdorff distance). *Let X be a metric space with metric d, $A, B \subset X$, $A, B \neq \emptyset$.*

(i) The distance *of a point $x \in X$ to A is given by*

$$d(x, A) := \inf_{a \in A} d(x, a).$$

(ii) The gap *between A and B is defined as*

$$D_d(A, B) := \inf_{a \in A} d(a, B).$$

(iii) The excess *of A over B is defined as*

$$e_d(A, B) := \sup_{a \in A} d(a, B).$$

(iv) The Hausdorff distance *is then defined as*

$$H_d(A, B) := \max\{e_d(A, B), e_d(B, A)\}.$$

Figure B.1 illustrates these definitions in the case of both intersecting and disjoint sets.

Remark B.2.

- *Note that the gap between the sets A and B is symmetric, i.e. $D_d(A, B) = D_d(B, A)$, in contrast to the excess, $e_d(A, B) \neq e_d(B, A)$.*

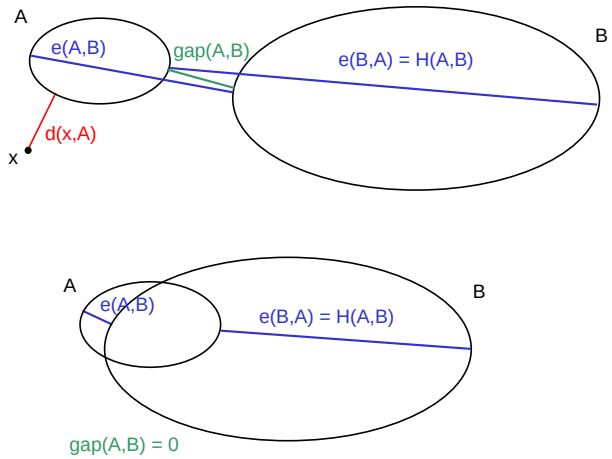


Figure B.1: Illustration of the Hausdorff distance and associated definitions.

- *For A, B compact, the gap $D_d(A, B)$ is equal to zero if the sets A and B have nonempty intersection, otherwise $D_d(A, B) > 0$.*
- *For A, B compact, the excess $e_d(A, B)$ is equal to zero if $A \subseteq B$.*
- *For A, B compact, the Hausdorff distance $H_d(A, B)$ is only equal to zero if the two sets are identical, otherwise $H_d(A, B) > 0$.*

Hausdorff lower and upper semicontinuity can be characterized as well using the definition of the excess of one set over another.

Proposition B.3. *Consider a set-valued mapping $\Gamma : \mathcal{U} \to \mathcal{P}(\mathbb{R}^n)$ and let $\{u_k\} \subset \mathcal{U}$ be a sequence with $u_k \to \hat{u}$. Then the following statements hold:*

- *The mapping Γ is H-usc at $\hat{u}$ according to Definition 2.23 if and only if*

$$e_d(\Gamma(u_k), \Gamma(\hat{u})) \to 0 \quad \text{for } u_k \to \hat{u}.$$

- *The mapping Γ is H-lsc at $\hat{u}$ according to Definition 2.23 if and only if*

$$e_d(\Gamma(\hat{u}), \Gamma(u_k)) \to 0 \quad \text{for } u_k \to \hat{u}.$$

Proof.

- For the forward direction, let Γ be H-usc at $\hat{u}$. We thus get the following implications:

$$\begin{aligned}
& \forall \varepsilon > 0 \; \exists \delta > 0 : \Gamma(u_k) \subset V_{\frac{\varepsilon}{2}}(\Gamma(\hat{u})) \quad \forall u_k \in V_\delta(\hat{u}) \\
\Rightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 : d(v, \Gamma(\hat{u})) < \frac{\varepsilon}{2} \quad \forall v \in \Gamma(u_k) \quad \forall u_k \in V_\delta(\hat{u}) \\
\Rightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 : \sup_{v \in \Gamma(u_k)} d(v, \Gamma(\hat{u})) \leq \frac{\varepsilon}{2} \quad \forall u_k \in V_\delta(\hat{u}) \\
\Rightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 : e_d(\Gamma(u_k), \Gamma(\hat{u})) < \varepsilon \quad \forall u_k \in V_\delta(\hat{u}) \\
\Rightarrow \quad & e_d(\Gamma(u_k), \Gamma(\hat{u})) \to 0 \quad \text{for } u_k \to \hat{u}.
\end{aligned}$$

  The backward direction is proved by similar arguments, starting with

$$e_d(\Gamma(u_k), \Gamma(\hat{u})) \to 0 \quad \text{for } u_k \to \hat{u}.$$

  It then holds:

$$\begin{aligned}
\Rightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 : \\
& e_d(\Gamma(u_k), \Gamma(\hat{u})) = \sup_{v \in \Gamma(u_k)} d(v, \Gamma(\hat{u})) < \varepsilon \quad \forall u_k \in V_\delta(\hat{u}) \\
\Rightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 : d(v, \Gamma(\hat{u})) < \varepsilon \quad \forall v \in \Gamma(u_k) \quad \forall u_k \in V_\delta(\hat{u}) \\
\Rightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 : \Gamma(u_k) \subset V_\varepsilon(\Gamma(\hat{u})) \quad \forall u_k \in V_\delta(\hat{u}).
\end{aligned}$$

- To prove the forward direction, let Γ be H-lsc at $\hat{u}$. Then we get the following implications:

$$\begin{aligned}
& \forall \varepsilon > 0 \; \exists \delta > 0 : \Gamma(\hat{u}) \subset V_{\frac{\varepsilon}{2}}(\Gamma(u_k)) \quad \forall u_k \in V_\delta(\hat{u}) \\
\Rightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 : d(w, \Gamma(u_k)) < \frac{\varepsilon}{2} \quad \forall w \in \Gamma(\hat{u}) \quad \forall u_k \in V_\delta(\hat{u}) \\
\Rightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 : \sup_{w \in \Gamma(\hat{u})} d(w, \Gamma(u_k)) \leq \frac{\varepsilon}{2} \quad \forall u_k \in V_\delta(\hat{u}) \\
\Rightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 : e_d(\Gamma(\hat{u}), \Gamma(u_k)) < \varepsilon \quad \forall u_k \in V_\delta(\hat{u}) \\
\Rightarrow \quad & e_d(\Gamma(\hat{u}), \Gamma(u_k)) \to 0 \quad \text{for } u_k \to \hat{u}.
\end{aligned}$$

  Backwards, let $e_d(\Gamma(\hat{u}), \Gamma(u_k)) \to 0$ for $u_k \to \hat{u}$. Then we obtain the following:

$$\begin{aligned}
\Rightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 : \\
& e_d(\Gamma(\hat{u}), \Gamma(u_k)) = \sup_{w \in \Gamma(\hat{u})} d(w, \Gamma(u_k)) < \varepsilon \quad \forall u_k \in V_\delta(\hat{u}) \\
\Rightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 : d(w, \Gamma(u_k)) < \varepsilon \quad \forall w \in \Gamma(\hat{u}) \quad \forall u_k \in V_\delta(\hat{u}) \\
\Rightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 : \Gamma(\hat{u}) \subset V_\varepsilon(\Gamma(u_k)) \quad \forall u_k \in V_\delta(\hat{u}).
\end{aligned}$$

□

Appendix C

Matrix analysis

In this short appendix, we summarize some useful facts and calculation rules for the trace and the Kronecker product of matrices.

Definition C.1. *The* trace *of an $n \times n$ matrix $A = (a_{ij})$ is defined to be the sum of the diagonal entries of A, i.e.*

$$tr(A) = \sum_{i=1}^{n} a_{ii}.$$

A rather useful property of the trace of a matrix product is that is remains unaffected when the order of the matrices is changed cyclically:

Lemma C.2.

(i) Let $A \in \mathbb{R}^{n\times k}$ and $B \in \mathbb{R}^{k\times n}$ be arbitrary matrices. It holds that

$$tr(AB) = tr(BA).$$

(ii) Let A, B and C be suitably sized matrices such that the following matrix products are possible. Then it holds that

$$tr(ABC) = tr(BCA) = tr(CAB). \tag{C.1}$$

For a proof see e.g. Meyer [59], page 110.

As a consequence of this lemma, we also obtain the following chain of equalities for $x \in \mathbb{R}^n$ and $C \in \mathbb{R}^{n\times n}$:

$$x^T C x = \mathrm{tr}(x^T C x) = \mathrm{tr}(C x x^T) = \mathrm{tr}(x x^T C).$$

Using the inner product on the space of symmetric matrices, an appropriate norm can be defined.

Definition C.3. *Let $A \in S^n$, i.e. A is a symmetric $n \times n$ matrix. The (trace-) norm of A, denoted by $\|A\|_{tr}$ is defined by*

$$\|A\|_{tr}^2 := \langle A, A \rangle = tr(A^T A) = tr(A^2).$$

The link between the trace-norm for matrices and the Euclidean norm for vectors is straightforward if a matrix $A \in S^n$ is restructured and interpreted as a vector in $\mathbb{R}^{n^2}$ by stacking the columns of A successively underneath each other. A vector transformed in such a way is denoted by $\text{vec}(A)$. For $A \in S^n$ we then obtain, using Equation (C.12) from below, that

$$\|A\|_{\text{tr}}^2 = \text{tr}(AA) = \text{vec}(A^T)^T \,\text{vec}(A) = \|\,\text{vec}(A)\|_2^2. \tag{C.2}$$

With help of the trace-norm, we can furthermore define the distance between pairs that consist of a vector and a symmetric matrix, i.e. given by (a, A) with $a \in R^n$ and $A \in S^n$.

Definition C.4. *Let pairs (a, A) and (b, B) with $a, b \in R^n$ and $A, B \in S^n$ be given. The distance between (a, A) and (b, B) is then defined as*

$$d\big((a, A), (b, B)\big) := \|a - b\|_2 + \|A - B\|_{tr}. \tag{C.3}$$

Note that this formula would also be obtained if a pair (a, A) was interpreted as a vector $w \in \mathbb{R}^{n+n^2}$ with $w = \begin{pmatrix} a \\ \text{vec}(A) \end{pmatrix}$ and the usual Euclidean distance was applied.

The transformation of a matrix into a vector is not only useful when dealing with norms, but as well for determining distributional properties of a random matrix. In the general case of an arbitrary $n \times k$ matrix A, the resulting vector $\text{vec}(A)$ has nk components and is also obtained by stacking the columns of A successively underneath each other. The relationship between $\text{vec}(A)$ and $\text{vec}(A^T)$ is described through the commutation matrix K_{nk} which is uniquely defined by the following equation:

$$\text{vec}(A) = K_{nk}\,\text{vec}(A^T) \quad \forall A \in \mathbb{R}^{n \times k}. \tag{C.4}$$

For the commutation matrix K_{nk}, we obtain the following results.

Lemma C.5.

(i) For a commutation matrix K_{nk} it holds that

$$K_{nk}^T = K_{nk}^{-1} = K_{kn}.$$

(ii) In case of quadratic matrices this simplifies further to

$$K_{nn}^T = K_{nn}^{-1} = K_{nn}. \tag{C.5}$$

Proof. See e.g. Meucci [57], Appendix A.6. □

Definition C.6. *The* Kronecker product *for two arbitrary matrices $A \in \mathbb{R}^{n\times k}$ and $B \in \mathbb{R}^{p\times q}$ is given by the $np \times kq$ matrix*

$$A \otimes B = \begin{pmatrix} A_{11}B & \cdots & A_{1k}B \\ \vdots & \ddots & \vdots \\ A_{n1}B & \cdots & A_{nk}B \end{pmatrix}.$$

For calculations including the Kronecker product and the vector notation of matrices, the following results are useful.

Lemma C.7. *Let A, B, C and D be suitably sized matrices such that the occurring products exist.*

- *For the Kronecker product, the following calculation rules apply:*

$$(A \otimes B) \cdot (C \otimes D) = (AC \otimes BD), \tag{C.6}$$
$$(A \otimes B)^{-1} = (A^{-1} \otimes B^{-1}), \tag{C.7}$$
$$(A \otimes B)^T = (A^T \otimes B^T), \tag{C.8}$$
$$A \otimes (B + C) = (A \otimes B) + (A \otimes C), \tag{C.9}$$
$$(A + B) \otimes C = (A \otimes C) + (B \otimes C). \tag{C.10}$$

- *In combination with the trace or the vector notation* vec(.)*, it holds:*

$$\mathrm{vec}(ABC) = (C^T \otimes A)vec(B), \tag{C.11}$$
$$tr(AB) = \mathrm{vec}(A^T)^T \,\mathrm{vec}(B). \tag{C.12}$$

- *For $E \in \mathbb{R}^{m\times m}$ and $F \in \mathbb{R}^{n\times n}$, we have:*

$$(E \otimes I_n)(I_m \otimes F) = (E \otimes F) = (I_m \otimes F)(E \otimes I_n). \tag{C.13}$$

Proof. See Meucci [57], Appendix A.6 and Meyer [59], pages 380 and 598. □

Appendix D

Selected distributions

In this appendix we summarize some selected elliptical distributions and their properties which are mainly applied in the Bayesian approach to determine parameter estimates in Chapter 7.

The probably most well-known and widely used representative of elliptical distributions is the normal distribution. Further, we will shortly introduce the Student-t distribution as well, which allows to model heavier tails than the normal distribution and which tends to a normal distribution for an increasing number of degrees of freedom.

Fang and Zhang [30] extend the analysis of elliptical distribution for random vectors also to random matrices. A special elliptical matrix distribution is the Wishart distribution which is used in asset management to discribe the distribution of a covariance matrix. We will furthermore present the inverse Wishart distribution and the normal inverse Wishart distribution, a combination of a normal and an inverse Wishart distribution which is used to model the joint behavior of a pair consisting of a random vector and a random matrix, e.g. the mean vector and the covariance matrix of asset returns.

D.1 Multivariate normal distribution

We will define the multivariate normal distribution straightforwardly through its density function. Equivalently, it can be defined in terms of a suitable linear transformation of a vector of (univariate) standard normally distributed variables.

Definition D.1. *A random variable $X \in \mathbb{R}^n$ with mean vector μ and covariance matrix Σ is said to be (multivariate) normally distributed, denoted by $X \sim \mathcal{N}(\mu, \Sigma)$, if the probability density function is given by*

$$\varphi_{\mathcal{N}}(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} |\Sigma|^{-\frac{1}{2}} \exp\left\{ \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) \right\}.$$

A rather useful result concerns the conditional distribution of a partitioned multivariate normally distributed random variable. The proof of this proposition can be found in various references, see e.g. Press [65], Theorem 3.5.1, or Fang and Zhang [30], Theorem 2.3.5.

Proposition D.2. *Let $X \sim \mathcal{N}(\mu, \Sigma)$ and partition $X \in \mathbb{R}^n$, $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$ as follows – with appropriate dimensions k and $n-k$:*

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Then, the conditioning formula *for the normal distribution, i.e. the conditional distribution of X_1 given $X_2 = x_2$ is expressed by*

$$(X_1 | X_2 = x_2) \sim \mathcal{N}(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}). \tag{D.1}$$

This conditioning formula for the normal distribution is more widely known than the general formulation for arbitrary elliptical distributions as given in Proposition 4.10, part (iv). This Equation (D.1) is needed in Chapter 7 to calculate the Bayesian posterior distribution.

D.2 Student-t distribution

A further elliptical distribution is the Student-t distribution which in the one-dimensional case is defined in terms of a normal and a χ^2_r-distribution: Let $Y \in \mathbb{R}, Y \sim \mathcal{N}(0,1)$ be independent from $Z \in \mathbb{R}, Z \sim \chi^2_r$, then $X = \frac{Y}{\sqrt{\frac{Z}{r}}}$ is Student-t distributed with r degrees of freedom. For the multivariate case we will use the following definition.

Definition D.3. *A random variable $X \in \mathbb{R}^n$ is said to be (multivariate) Student-t distributed with r degrees of freedom and parameters μ and Σ, denoted by $X \sim St(\mu, \Sigma, r)$, if its density function has the form (see Meucci, Formula (2.188))*

$$\varphi_{St}(x) = (r\pi)^{-\frac{n}{2}} \frac{\Gamma\left(\frac{r+n}{2}\right)}{\Gamma\left(\frac{r}{2}\right)} |\Sigma|^{-\frac{1}{2}} \cdot \left(1 + \frac{1}{r}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)^{-\frac{r+n}{2}}.$$

Besides the density function of the Student-t distribution, we will also need the moments.

Proposition D.4. *Let $X \sim St(\mu, \Sigma, r)$. The moments of X are thus given by*

$$\mathbf{E}[X] = \mu, \quad \textit{and} \quad \mathbf{Cov}[X] = \frac{r}{r-2} \cdot \Sigma.$$

Proof. See e.g. Press [65], page 128, or Meucci [57], page 79. □

Unlike the normal distribution, the Student-t distribution can be used to model fat tails. As the number of degrees of freedom becomes large ($r > 30$) the probability density function of the Student-t distribution with r degrees of freedom resembles more and more the probability density function of the normal distribution.

D.3 Wishart distribution and Wishart related distributions

The Wishart distribution is a distribution for symmetric and positive definite matrices in $\mathbb{R}^{n\times n}$. It represents a generalization of the (one-dimensional) χ^2_ν-distribution, i.e. in the case of $n = 1$ the Wishart probability density function also reduces to the pdf of a χ^2_ν-disribution.

Definition D.5. *Let $\Sigma \in \mathbb{R}^{n\times n}$ be a symmetric and positive definite matrix. The random variable Σ is said to be Wishart distributed with scale matrix C and ν degrees of freedom, denoted by $\Sigma \sim \mathcal{W}(C,\nu)$, if the density function is given by the formula*

$$\varphi_{\mathcal{W}}(\Sigma) = \begin{cases} \frac{c_{\mathcal{W}}|\Sigma|^{\frac{\nu-n-1}{2}}}{|C|^{\frac{\nu}{2}}} \exp\left(-\frac{1}{2}tr(C^{-1}\Sigma)\right) & \text{if } \Sigma \succ 0, \Sigma = \Sigma^T, \\ 0 & \text{otherwise,} \end{cases}$$

with

$$c_{\mathcal{W}} = \left[2^{\frac{\nu n}{2}} \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\left(\frac{\nu+1-i}{2}\right)\right]^{-1}$$

being a normalizing constant and Γ denoting the Gamma function.

Proposition D.6. *Let $\Sigma \sim \mathcal{W}(C,\nu)$. Then it holds*

$$\begin{aligned} \mathbf{E}[\Sigma] &= \nu \cdot C, \\ \mathbf{Var}[\Sigma_{ij}] &= \nu \cdot \left(C_{ij}^2 + C_{ii} \cdot C_{jj}\right), \\ \mathbf{Cov}[\Sigma_{ij}, \Sigma_{kl}] &= \nu \cdot \left(C_{ik} \cdot C_{jl} + C_{il} \cdot C_{jk}\right). \end{aligned}$$

Proof. See Press [65], Theorem 5.1.7. □

Remark D.7. *In asset management practice the case where $\Sigma \sim \mathcal{W}(\frac{1}{\nu}C,\nu)$ is rather common, as this represents the distribution of the empirical covariance matrix estimator based on a sample of normally distributed random vectors. Then*

the above formulas become

$$\begin{aligned}\mathbf{E}[\Sigma] &= \nu \cdot \frac{1}{\nu} C = C,\\ \mathbf{Var}[\Sigma_{ij}] &= \frac{1}{\nu}\left(C_{ij}^2 + C_{ii}C_{jj}\right),\\ \mathbf{Cov}[\Sigma_{ij}, \Sigma_{kl}] &= \frac{1}{\nu}\left(C_{ik}C_{jl} + C_{il}C_{jk}\right).\end{aligned}$$

The following *inverse Wishart* distribution often occurs when the joint distribution of a vector and a matrix is needed. We will see explicit calculations with this inverse Wishart distribution later on in the definition of the *normal inverse Wishart* distribution. Additionally, these types of distributions will be used extensively in Section 7.1 in the Bayesian approach to determine parameters for the portfolio optimization problem.

Definition D.8. *Let $\Sigma \in \mathbb{R}^{n\times n}, \Sigma \sim \mathcal{W}(C,\nu)$. Then Σ^{-1} follows an inverse Wishart distribution with scale matrix C^{-1} and $\nu + n + 1$ degrees of freedom. This will be denoted by $\Sigma^{-1} \sim \mathcal{IW}(C^{-1}, \nu + n + 1)$.*

For ease of notation, in the following proposition we will express the moments and the density of the inverse Wishart distribution not in terms of the inverse of a matrix, but simply in terms of a matrix itself, i.e. as $U \sim \mathcal{IW}(V,k)$. As the moments and the density of the inverse Wishart distribution, given in the subsequent proposition, are only defined[1] for $k > 2n + 4$, we make the next assumption.

Assumption D.9. *In the following, we will always assume that the number of degrees of freedom in the inverse Wishart distribution is sufficiently large to meet all technical requirements.*

Remark D.10. *Assumption D.9 is not a very strong assumption in practice. It is known (see e.g. Press [65], Theorem 7.1.5) that the maximum likelihood estimator for the covariance matrix based on a sample of size S follows a Wishart distribution with $S-1$ degrees of freedom. Thus, for the moments and the density of the inverse matrix to be defined, it is required that the number of degrees of freedom of the inverse Wishart distribution, $k = (S-1) + n + 1 = S + n$, is larger than $2n + 4$. Hence, a sample of size S with $S > n + 4$ suffices to assure existence of the moments and the density of the inverse Wishart distribution. Since it is anyway necessary to have a sample size of at least n to guarantee positive definiteness of the covariance matrix, this assumption is usually fulfilled in practice.*

[1] To be precise, for the first moment it is sufficient to have $k > 2n + 2$, the second moment requires $k > 2n + 4$, and the density function is already defined for $k > 2n$, see Press [65], page 110 and Theorem 5.2.2, respectively.

Now we are ready to state the following proposition.

Proposition D.11. *Let U be a symmetric and positive definite matrix, and let $U \sim \mathcal{IW}(V, k)$. The moments of the inverse Wishart distributed variable U are given by*

$$\begin{aligned}
\mathbf{E}[U] &= \frac{V}{k-2n-2}, \\
\mathbf{Var}[U_{ii}] &= \frac{2(V_{ii})^2}{(k-2n-2)^2(k-2n-4)}, \\
\mathbf{Var}[U_{ij}] &= \frac{V_{ii}V_{jj} + \frac{k-2n}{k-2n-2}(V_{ij})^2}{(k-2n-1)(k-2n-2)(k-2n-4)}, \\
\mathbf{Cov}[U_{ij}, U_{kl}] &= \frac{\frac{2}{k-2n-2}V_{ij}V_{kl} + V_{ik}V_{jl} + V_{il}V_{kj}}{(k-2n-1)(k-2n-2)(k-2n-4)}.
\end{aligned}$$

The density of the inverse Wishart distribution is given by

$$\varphi_{\mathcal{IW}}(U) = \begin{cases} \frac{c_{\mathcal{IW}}|V|^{\frac{k-n-1}{2}}}{|U|^{\frac{k}{2}}} \exp\left(-\frac{1}{2}tr(U^{-1}V)\right) & \text{if } U \succ 0, U = U^T, \\ 0 & \text{otherwise,} \end{cases}$$

with

$$c_{\mathcal{IW}} = \left[2^{\frac{(k-n-1)n}{2}} \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\left(\frac{k-n-i}{2}\right)\right]^{-1}$$

being an appropriate constant.

Proof. See Press [65], page 110 and Theorem 5.2.2. □

Remark D.12. *Vice versa, having $U \sim \mathcal{IW}(C^{-1}, k)$, we obtain that*

$$\Sigma := U^{-1} \sim \mathcal{W}(C, k-n-1).$$

After having defined a matrix distribution, we want to combine the normal and the inverse Wishart distribution to describe the joint behavior of a vector and a matrix. In the literature this is sometimes expressed in terms of a parameter pair (μ, Σ) and sometimes in terms of (μ, Σ^{-1}). This is merely notational convention and does not affect the results obtained thereof. We will use (μ, Σ) to define the normal inverse Wishart distribution and thus as well to perform any further calculations.

Definition D.13. *The pair (μ, Σ) is said to be distributed according to a normal inverse Wishart distribution with parameters μ_0, d_0, Σ_0, and ν_0, denoted by $(\mu, \Sigma) \sim \mathcal{NIW}(\mu_0, d_0, \Sigma_0, \nu_0)$, if it holds that*

$$\mu \mid \Sigma \sim \mathcal{N}\left(\mu_0, \frac{1}{d_0}\Sigma\right) \quad \text{and} \quad \Sigma \sim \mathcal{IW}(\nu_0\Sigma_0, \nu_0 + n + 1).$$

From this definition the moments of $\mu \mid \Sigma$ and Σ are readily given by the according propositions above. The joint density function of (μ, Σ) can straightforwardly be calculated by the Bayes rule and the individual density functions.

Proposition D.14 (Joint density function).
Let $(\mu, \Sigma) \sim \mathcal{NIW}(\mu_0, d_0, \Sigma_0, \nu_0)$. The joint density function is then given by

$$\begin{aligned}\varphi_{\mathcal{NIW}}(\mu, \Sigma) &= \\ &= \gamma \ |\Sigma^{-1}|^{\frac{\nu_0+n+2}{2}} \exp\left\{-\frac{1}{2}[d_0(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0) + tr(\nu_0\Sigma_0 \cdot \Sigma^{-1})]\right\} \quad \text{(D.2)}\end{aligned}$$

with $\gamma = \frac{1}{(2\pi)^{\frac{n}{2}}} \ c_{\mathcal{IW}} \ |r_0\Sigma_0|^{\frac{\nu_0}{2}}$.

Proof. Follows directly from the Bayes rule for conditional distributions,

$$\varphi_{\mathcal{NIW}}(\mu, \Sigma) = \varphi_{\mathcal{N}}(\mu \mid \Sigma) \ \varphi_{\mathcal{IW}}(\Sigma). \qquad \square$$

As mentioned above, from the definition of the normal inverse Wishart distribution, the moments of Σ are already given. The moments of μ itself, i.e. not conditioned on the matrix Σ, are not available directly thereof, the marginal distribution of μ is required beforehand.

Proposition D.15 (Marginal distribution of μ).
Let $(\mu, \Sigma) \sim \mathcal{NIW}(\mu_0, d_0, \Sigma_0, \nu_0)$. Then it holds that

$$\mu \sim St\left(\mu_0, \frac{\nu_0}{\nu_0 - n + 1} \cdot \frac{\Sigma_0}{d_0}, \nu_0 - n + 1\right)$$

with the associated moments

$$\begin{aligned}\mathbf{E}[\mu] &= \mu_0, \\ \mathbf{Cov}[\mu] &= \frac{\nu_0}{\nu_0 - n - 1} \cdot \frac{\Sigma_0}{d_0}.\end{aligned}$$

Proof. This proof will show the explicit calculations using the notation in terms of (μ, Σ), closely following the calculations with (μ, Σ^{-1}) which are given in Meucci [57], Appendix www.7.5. The expressions α_1, α_2, α_3 and α_4 are supposed to be normalizing constants such that the respective formulas represent probability density functions.

With the definition

$$\Sigma_2 := d_0(\mu - \mu_0)(\mu - \mu_0)^T + \nu_0\Sigma_0$$

the joint distribution of μ and Σ can be written as

$$\varphi_{\mathcal{NIW}}(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}} c_{\mathcal{IW}} |\nu_0\Sigma_0|^{\frac{\nu_0}{2}} |\Sigma^{-1}|^{\frac{\nu_0+n+2}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(\Sigma_2\Sigma^{-1})\right\}.$$

Thus, the marginal distribution of μ is obtained by

$$\begin{aligned}
\varphi(\mu) &= \int \varphi_{\mathcal{NIW}}(\mu, \Sigma)\, d\Sigma \\
&= \alpha_1 |\nu_0 \Sigma_0|^{\frac{\nu_0}{2}} \int |\Sigma|^{-\frac{\nu_0+n+2}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(\Sigma_2 \Sigma^{-1})\right\} d\Sigma \\
&= \alpha_1 |\nu_0 \Sigma_0|^{\frac{\nu_0}{2}} |\Sigma_2|^{-\frac{\nu_0+1}{2}} \frac{1}{\tilde{c}_{\mathcal{IW}}} \cdot \\
&\quad \cdot \int \tilde{c}_{\mathcal{IW}} |\Sigma_2|^{\frac{\nu_0+1}{2}} |\Sigma|^{-\frac{\nu_0+n+2}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(\Sigma_2 \Sigma^{-1})\right\} d\Sigma
\end{aligned}$$

where in the last equation we have completed the integrand to represent the density of Σ with $\Sigma \sim \mathcal{IW}(\Sigma_2, \nu_0 + n + 2)$. Thus, the entire integral is equal to 1. Subsuming the constant $\frac{1}{\tilde{c}_{\mathcal{IW}}}$ into the normalizing value α_2, we continue with

$$\begin{aligned}
\varphi(\mu) &= \alpha_2 |\nu_0 \Sigma_0|^{\frac{\nu_0}{2}} |\Sigma_2|^{-\frac{\nu_0+1}{2}} \\
&= \alpha_2 |\nu_0 \Sigma_0|^{\frac{\nu_0}{2}} |\nu_0 \Sigma_0 + d_0 (\mu - \mu_0)(\mu - \mu_0)^T|^{-\frac{\nu_0+1}{2}}
\end{aligned}$$

which can be reformulated using the equation

$$|A|^{\frac{\nu_0}{2}} \cdot |A + vv^T|^{-\frac{\nu_0+1}{2}} = |A|^{-\frac{1}{2}} (1 + v^T A^{-1} v)^{-\frac{\nu_0+1}{2}},$$

see e.g. Meucci [57], Equation (T7.82) in Appendix www.7.5, and thus gives

$$\begin{aligned}
\varphi(\mu) &= \alpha_2 |\nu_0 \Sigma_0|^{-\frac{1}{2}} \left(1 + d_0 (\mu - \mu_0)^T (\nu_0 \Sigma_0)^{-1} (\mu - \mu_0)\right)^{-\frac{\nu_0+1}{2}} \\
&= \alpha_2 \nu_0^{-\frac{n}{2}} d_0^{-\frac{n}{2}} \frac{1}{d_0^{-\frac{n}{2}}} |\Sigma_0|^{-\frac{1}{2}} \left(1 + \left(\frac{1}{d_0}\right)^{-1} \nu_0^{-1} (\mu - \mu_0)^T \Sigma_0^{-1} (\mu - \mu_0)\right)^{-\frac{\nu_0+1}{2}} \\
&= \alpha_3 \left|\frac{\Sigma_0}{d_0}\right|^{-\frac{1}{2}} \left(1 + \frac{1}{\nu_0} (\mu - \mu_0)^T \left(\frac{\Sigma_0}{d_0}\right)^{-1} (\mu - \mu_0)\right)^{-\frac{\nu_0+1}{2}}.
\end{aligned}$$

Substituting $z := \nu_0 - n + 1$, this can be transformed to

$$\begin{aligned}
\varphi(\mu) &= \alpha_3 \left|\frac{\Sigma_0}{d_0}\right|^{-\frac{1}{2}} \cdot \left(\frac{z+n-1}{z+n-1} \cdot \frac{z}{z}\right)^{-\frac{n}{2}} \\
&\quad \cdot \left(1 + \frac{1}{z+n-1} \cdot \frac{z}{z} \cdot (\mu - \mu_0)^T \left(\frac{\Sigma_0}{d_0}\right)^{-1} (\mu - \mu_0)\right)^{-\frac{z+n}{2}} \\
&= \alpha_4 \left|\frac{z+n-1}{z} \cdot \frac{\Sigma_0}{d_0}\right|^{-\frac{1}{2}} \left(1 + \frac{1}{z} (\mu - \mu_0)^T \left(\frac{z+n-1}{z} \cdot \frac{\Sigma_0}{d_0}\right)^{-1} (\mu - \mu_0)\right)^{-\frac{z+n}{2}}.
\end{aligned}$$

Thus, comparing this last equation with the formula given in Definition D.3, this is the density of the following Student-t distribution:

$$\mu \sim St\left(\mu_0, \frac{z+n-1}{z} \cdot \frac{\Sigma_0}{d_0}, z\right)$$

or equivalently, expressed again in terms of ν_0

$$\mu \sim St\left(\mu_0, \frac{\nu_0}{\nu_0 - n + 1} \cdot \frac{\Sigma_0}{d_0}, \nu_0 - n + 1\right)$$

which implies the first two moments

$$\begin{aligned} \mathbf{E}[\mu] &= \mu_0, \\ \mathbf{Cov}[\mu] &= \frac{\nu_0 - n + 1}{(\nu_0 - n + 1) - 2} \cdot \frac{\nu_0}{\nu_0 - n + 1} \cdot \frac{\Sigma_0}{d_0} \\ &= \frac{\nu_0}{\nu_0 - n - 1} \cdot \frac{\Sigma_0}{d_0}. \end{aligned}$$

□

Appendix E

Equivalent representations of an ellipsoidal uncertainty set

We want to show in this short appendix that the representations for an ellipsoidal uncertainty set used in Example 3.26 are in fact equivalent.

Lemma E.1. *Let $\hat{u} \in \mathcal{U}$, $\delta > 0$ and let the matrix Σ be symmetric and positive definite. Then it holds that*

$$\left\{u \in \mathcal{U} \mid (u-\hat{u})^T \Sigma^{-1} (u-\hat{u}) \leq \delta^2\right\} = \left\{u \in \mathcal{U} \mid u = \hat{u} + \delta\, \Sigma^{\frac{1}{2}} w, \|w\| \leq 1\right\}.$$

Proof. We prove equivalence of these two sets by showing that each side is a subset of the other one.

- Let u be such that $(u-\hat{u})^T \Sigma^{-1} (u-\hat{u}) \leq \delta^2$ holds. Since Σ is symmetric and positive definite, there exists the square root matrix $\Sigma^{\frac{1}{2}}$ which is as well symmetric and positive definite. We can thus make the equivalent reformulations

$$\begin{aligned} & (u-\hat{u})^T \Sigma^{-1} (u-\hat{u}) \leq \delta^2 \\ \Leftrightarrow \quad & (u-\hat{u})^T \Sigma^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} (u-\hat{u}) \leq \delta^2 \\ \Leftrightarrow \quad & \left(\Sigma^{-\frac{1}{2}}(u-\hat{u})\right)^T \left(\Sigma^{-\frac{1}{2}}(u-\hat{u})\right) \leq \delta^2. \end{aligned}$$

  Defining $w := \frac{1}{\delta} \Sigma^{-\frac{1}{2}} (u-\hat{u})$ yields

$$u = \hat{u} + \delta \Sigma^{\frac{1}{2}} w,$$

  and the above inequality reduces to $w^T w = \|w\|^2 \leq 1$, thus equivalent to $\|w\| \leq 1$.

- Let u be such that $u = \hat{u} + \delta \Sigma^{\frac{1}{2}} w$ holds with $\|w\| \leq 1$. Then we get

$$\begin{aligned}(u-\hat{u})^T\Sigma^{-1}(u-\hat{u}) &= (\delta\Sigma^{\frac{1}{2}}w)^T\Sigma^{-1}(\delta\Sigma^{\frac{1}{2}}w)\\ &= \delta^2(\Sigma^{\frac{1}{2}}w)^T\Sigma^{-1}(\Sigma^{\frac{1}{2}}w)\\ &= \delta^2 w^T w\\ &= \delta^2\|w\|^2\\ &\leq \delta^2.\end{aligned}$$

□

Appendix F

Reformulation of (GCP_u) and $(LRC_{\hat{u},\delta})$

In this appendix we want to show that we can assume without loss of generality that the objective function is linear in x and independent of u and that this assumption remains valid when applying the local robust counterpart approach.

Consider the following two programs

$$\begin{aligned} \min_{x \in X} \quad & f(x,u) \\ \text{s.t.} \quad & g(x,u) \leq_K 0 \end{aligned} \tag{P_u}$$

and

$$\begin{aligned} \min_{\tilde{x} \in \tilde{X}} \quad & l(\tilde{x}) \\ \text{s.t.} \quad & \tilde{g}(\tilde{x},u) \leq_{\tilde{K}} 0 \end{aligned} \tag{$\tilde{P}_u$}$$

where

- $X \subseteq \mathbb{R}^n$ is non-empty, convex and compact,
- $\mathcal{U} \subset \mathbb{R}^d$ is non-empty, convex and compact,
- $z \in Z = [Z_l, Z_u] \subset \mathbb{R}$ with

$$\begin{aligned} Z_l &= \min_{x \in X} \min_{u \in \mathcal{U}} f(x,u) - 1 < \infty \\ Z_u &= \max_{x \in X} \max_{u \in \mathcal{U}} f(x,u) + 1 < \infty, \end{aligned}$$

- $\tilde{X} = X \times Z$ is a compact set,
- $\tilde{x} = (x_1, \ldots, x_n, z)^T \in \tilde{X}$,

- $l : \mathbb{R}^{n+1} \to \mathbb{R}$ with $l(\tilde{x}) = l(x_1, \ldots, x_n, z) = z$ is linear and continuous in $\tilde{x}$,
- $\tilde{g} : \mathbb{R}^{n+1} \times \mathbb{R}^d \to \mathbb{R}^{m+1}$ with

$$\tilde{g}(\tilde{x}, u) := \begin{pmatrix} g(x,u) \\ f(x,u) - z \end{pmatrix}$$

  is continuous both in $\tilde{x}$ and u,
- $\tilde{K} = K \times \mathbb{R}_+ \subset \mathbb{R}^{m+1}$,
- the relation $\leq_{\tilde{K}}$ is defined for $a, b \in \tilde{K} \subset \mathbb{R}^{m+1}$ as follows:

$$a \leq_{\tilde{K}} b \quad \Leftrightarrow \quad \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} \leq_K \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \quad \text{and} \quad a_{m+1} \leq b_{m+1}.$$

Note that with this definition of the relation $\leq_{\tilde{K}}$, K-convexity of $g(x,.)$ (resp. $g(.,u)$) and convexity of $f(x,.)$ (resp. $f(.,u)$) yield $\tilde{K}$-convexity of $\tilde{g}(x,.)$ (resp. $\tilde{g}(.,u)$), hence convexity of the program is maintained. Furthermore, the property of the existence of a Slater point remains as there exists a $z \in Z = [Z_l, Z_u]$ such that $f(x,u) < z$.

Before showing equivalence of the problems (P_u) and $(\tilde{P}_u)$, we prove that Lipschitz continuity of the individual functions f and g also transfers to the combined function $\tilde{g}$.

Proposition F.1. *Let f and g be globally Lipschitz continuous in u. Then $\tilde{g}$ is globally Lipschitz continuous in u as well.*

Proof. Let f and g be globally Lipschitz continuous in u with Lipschitz constants L_f and L_g, respectively, and let $u_1, u_2 \in \mathcal{U}$. It then holds that

$$\begin{aligned} \|\tilde{g}(\tilde{x}, u_1) - \tilde{g}(\tilde{x}, u_2)\|^2 &= \left\| \begin{pmatrix} g(x,u_1) - g(x,u_2) \\ \big(f(x,u_1) - z\big) - \big(f(x,u_2) - z\big) \end{pmatrix} \right\|^2 \\ &= \|g(x,u_1) - g(x,u_2)\|^2 + \|f(x,u_1) - f(x,u_2)\|^2 \\ &\leq \big(L_g \|u_1 - u_2\|\big)^2 + \big(L_f \|u_1 - u_2\|\big)^2 \\ &= \underbrace{\big(L_g^2 + L_f^2\big)}_{=:L^2} \|u_1 - u_2\|^2 \end{aligned}$$

and hence

$$\|\tilde{g}(\tilde{x}, u_1) - \tilde{g}(\tilde{x}, u_2)\| \leq L \|u_1 - u_2\|,$$

i.e. $\tilde{g}$ is globally Lipschitz continuous in u. □

Notation F.2. *We will denote the feasibility sets of (P_u) and $(\tilde{P}_u)$ with $\mathcal{F}_P(u)$ and $\mathcal{F}_{\tilde{P}}(u)$, respectively, i.e.*

$$\mathcal{F}_P(u) = \{x \in X \mid g(x,u) \leq_K 0\},$$
$$\mathcal{F}_{\tilde{P}}(u) = \{\tilde{x} \in \tilde{X} \mid \tilde{g}(\tilde{x},u) \leq_{\tilde{K}} 0\}.$$

Futhermore, the optimal value functions will be denoted by $f_P^*(u)$ and $f_{\tilde{P}}^*(u)$, respectively.

Proposition F.3. *The programs (P_u) and $(\tilde{P}_u)$ as denoted above are equivalent in the sense that for given $u \in \mathcal{U}$ each optimal (feasible) solution x^* of (P_u) with associated objective value $f(x^*,u)$ can be expanded to an optimal (feasible) solution $\tilde{x}^* = \left(x^{*T}, f(x^*,u)\right)^T$ of $(\tilde{P}_u)$. Conversely, for each optimal (feasible) solution $\tilde{x}^* = (x_1^*, \ldots, x_n^*, z^*)^T$ of $(\tilde{P}_u)$, the projection onto X, x^*, is an optimal (feasible) solution of (P_u).*

Proof. We will first consider the feasible points and afterwards deal with optimality.

Let x^* be feasible for (P_u). Then $g(x^*,u) \leq_K 0$ and with $z^* := f(x^*,u)$ it holds that $\tilde{g}(\tilde{x}^*,u) \leq_{\tilde{K}} 0$ for $\tilde{x}^* = (x^{*T}, z^*)^T$. Thus, the extended point $\tilde{x}^*$ is feasible for $(\tilde{P}_u)$.
Conversely, let $\tilde{x}^* = (x^{*T}, z^*)^T$ be feasible for $(\tilde{P}_u)$. Then $\tilde{g}(\tilde{x}^*,u) \leq_{\tilde{K}} 0$, which especially implies $g(x^*,u) \leq_K 0$, and thus x^* as the projection of $\tilde{x}^*$ onto X is feasible for (P_u).

To show optimality, we need to verify equality of the optimal values. Let x_P^* be an optimal solution of (P_u) with associated optimal objective value $z_P^* := f(x_P^*,u)$. Thus we have that $\tilde{x}_P^* := ({x_P^*}^T, z_P^*)^T$ is feasible for $(\tilde{P}_u)$ with objective value $l(\tilde{x}_P^*) = z_P^*$. Thus, we get

$$\begin{aligned} f_P^*(u) &= \min_{x \in \mathcal{F}_P(u)} f(x,u) \\ &= f(x_P^*,u) \\ &= z_P^* \\ &= l(\tilde{x}_P^*) \\ &\geq \min_{\tilde{x} \in \mathcal{F}_{\tilde{P}}(u)} l(\tilde{x}) = f_{\tilde{P}}^*(u). \end{aligned}$$

Conversely, let $\tilde{x}_{\tilde{P}}^* = ({x_{\tilde{P}}^*}^T, z_{\tilde{P}}^*)^T$ be an optimal solution of $(\tilde{P}_u)$ with associated optimal objective value $l(\tilde{x}_{\tilde{P}}^*) = z_{\tilde{P}}^*$. We know that $x_{\tilde{P}}^*$ is then feasible for (P_u)

and altogether we get

$$\begin{aligned} f^*_{\tilde{P}}(u) &= \min_{\tilde{x}\in\mathcal{F}_{\tilde{P}}(u)} l(\tilde{x}) \\ &= l(\tilde{x}^*_{\tilde{P}}) \\ &= z^*_{\tilde{P}} \\ &\geq f(x^*_{\tilde{P}}, u) \\ &\geq \min_{x\in\mathcal{F}_P(u)} f(x,u) = f^*_P(u). \end{aligned}$$

Thus, we have $f^*_P(u) = f^*_{\tilde{P}}(u)$ which then also implies from the above chains of equations that $\tilde{x}^*_P$ as the expanded optimal solution of (P_u) must be optimal for $(\tilde{P}_u)$ and analogously, $x^*_{\tilde{P}}$ as the reduced (i.e. projected onto X) optimal solution of $(\tilde{P}_u)$ must be optimal for (P_u). □

To illustrate that the simplifying assumption with respect to the objective function can also be made without loss of generality in case of the robust counterpart, we need to verify again that an optimal (feasible) solution of one program formulation can be expanded or reduced, respectively, to an optimal (feasible) solution of the other one. Thus, using the notation as introduced above, the two robust program formulations are the following:

$$\begin{aligned} \min_{x\in X} \quad & \max_{u\in\mathcal{U}} f(x,u) \\ \text{s.t.} \quad & g(x,u) \leq_K 0 \quad \forall u \in \mathcal{U} \end{aligned} \tag{R}$$

and

$$\begin{aligned} \min_{\tilde{x}\in\tilde{X}} \quad & \max_{u\in\mathcal{U}} l(\tilde{x}) \\ \text{s.t.} \quad & \tilde{g}(\tilde{x},u) \leq_{\tilde{K}} 0 \quad \forall u \in \mathcal{U} \end{aligned}$$

which is equivalent to

$$\begin{aligned} \min_{\tilde{x}\in\tilde{X}} \quad & l(\tilde{x}) \\ \text{s.t.} \quad & \tilde{g}(\tilde{x},u) \leq_{\tilde{K}} 0 \quad \forall u \in \mathcal{U}. \end{aligned} \tag{$\tilde{R}$}$$

Notation F.4. *We will denote the feasibility sets of (R) and ($\tilde{R}$) with $\mathcal{F}_R$ and $\mathcal{F}_{\tilde{R}}$, respectively, i.e.*

$$\begin{aligned} \mathcal{F}_R &= \{x \in X \mid g(x,u) \leq_K 0 \ \ \forall u \in \mathcal{U}\}, \\ \mathcal{F}_{\tilde{R}} &= \{\tilde{x} \in \tilde{X} \mid \tilde{g}(\tilde{x},u) \leq_{\tilde{K}} 0 \ \ \forall u \in \mathcal{U}\}. \end{aligned}$$

Accordingly, the optimal value functions will be denoted by f^*_R and $f^*_{\tilde{R}}$, respectively.

Proposition F.5. *The programs (R) and ($\tilde{R}$) as denoted above are equivalent in the sense that each optimal (feasible) solution x^* of (R) with associated objective value $\max_{u\in\mathcal{U}} f(x^*,u)$ can be expanded to an optimal (feasible) solution $\tilde{x}^* = \left(x^{*T}, \max_{u\in\mathcal{U}} f(x^*,u)\right)^T$ of ($\tilde{R}$). Conversely, for each optimal (feasible) solution $\tilde{x}^* = (x_1^*,\ldots,x_n^*,z^*)^T$ of ($\tilde{R}$), the projection onto X, x^*, is an optimal (feasible) solution of (R).*

Proof. We will again start with investigations on the feasibility.

Let x^* be feasible for (R). Then $g(x^*,u) \leq_K 0$ for all $u \in \mathcal{U}$ and with

$$z^* := \max_{u\in\mathcal{U}} \quad f(x^*,u)$$

it holds that $\tilde{g}(\tilde{x}^*,u) \leq_{\tilde{K}} 0$ for all $u \in \mathcal{U}$ with $\tilde{x}^* = (x^{*T}, z^*)^T$. Thus, the extended point $\tilde{x}^*$ is feasible for $(\tilde{R})$.
Conversely, let $\tilde{x}^* = (x^{*T}, z^*)^T$ be feasible for $(\tilde{R})$. Then $\tilde{g}(\tilde{x}^*,u) \leq_{\tilde{K}} 0$ for all $u \in \mathcal{U}$, which especially implies $g(x^*,u) \leq_K 0$ for all $u \in \mathcal{U}$ and thus x^* as the projection of $\tilde{x}^*$ onto X is feasible for (R).

Next we show equality of the optimal values. Let x_R^* be an optimal solution of (R) with associated optimal objective value

$$z_R^* := \max_{u\in\mathcal{U}} \quad f(x_R^*,u).$$

Thus, $\tilde{x}_R^* := (x_R^{*T}, z_R^*)^T$ is feasible for $(\tilde{R})$ with objective value $l(\tilde{x}_R^*) = z_R^*$ and we get

$$\begin{aligned} f_R^* &= \min_{x\in\mathcal{F}_R} \max_{u\in\mathcal{U}} f(x,u) \\ &= \max_{u\in\mathcal{U}} f(x_R^*,u) \\ &= z_R^* \\ &= l(\tilde{x}_R^*) \\ &\geq \min_{\tilde{x}\in\mathcal{F}_{\tilde{R}}} l(\tilde{x}) = f_{\tilde{R}}^*. \end{aligned}$$

Conversely, let $\tilde{x}_{\tilde{R}}^* = (x_{\tilde{R}}^{*T}, z_{\tilde{R}}^*)^T$ be an optimal solution of $(\tilde{R})$ with associated optimal objective value $l(\tilde{x}_{\tilde{R}}^*) = z_{\tilde{R}}^*$. We know that $x_{\tilde{R}}^*$ is then feasible for (R) and altogether we get

$$\begin{aligned} f_{\tilde{R}}^* &= \min_{\tilde{x}\in\mathcal{F}_{\tilde{R}}} l(\tilde{x}) \\ &= l(\tilde{x}_{\tilde{R}}^*) \\ &= z_{\tilde{R}}^* \\ &\geq f(x_{\tilde{R}}^*,u) \quad \forall u \in \mathcal{U}. \end{aligned}$$

Hence,

$$\begin{aligned} f^*_{\tilde{R}} &\geq \max_{u \in \mathcal{U}} f(x^*_{\tilde{R}}, u) \\ &\geq \min_{x \in \mathcal{F}_R} \max_{u \in \mathcal{U}} f(x, u) = f^*_R. \end{aligned}$$

Thus, we have $f^*_R = f^*_{\tilde{R}}$ which then also implies from the above chains of equations that $\tilde{x}^*_R$ as the expanded optimal solution of (R) must be optimal for $(\tilde{R})$ and analogously, $x^*_{\tilde{R}}$ as the reduced (i.e. projected onto X) optimal solution of $(\tilde{R})$ must be optimal for (R). □

Remark F.6. *As the maximum of convex functions is still convex, (K-)convexity of f and g yields $\tilde{K}$-convexity of $\tilde{g}$ (both in x and u). Furthermore, analogous to the original problem, the existence of a Slater point remains when modifying one problem into the other one.*

Appendix G

Detailed calculations to Example 3.27

This appendix shows the calculations that were performed to find the explicit formulas for the optimal solution and the associated optimal objective value in Example 3.27. Recall the optimization program:

$$\min_{x \in X} \quad -x^T u \tag{P}$$

with $X = \{x \in \mathbb{R}^2 \mid x \geq 0,\ x^T \mathbf{1} = 1\}$ and $u \in \mathcal{U}$.
Using the circular uncertainty set

$$\mathcal{U}_{\delta,\text{ell}}(\hat{u}) = \{u = \hat{u} + \delta w \mid \|w\|_2 \leq 1\}$$

the local robust counterpart problem is given by

$$\min_{x \in X} \max_{u \in \mathcal{U}_{\delta,\text{ell}}(\hat{u})} -x^T u \quad = \quad \min_{x \in X} -x^T \hat{u} + \delta \|x\|. \tag{LRC}$$

The Lagrangian function to this optimization problem is given for $\lambda \in \mathbb{R}$ by

$$L(x, \lambda) = -x^T \hat{u} + \delta \|x\|_2 + \lambda(x^T \mathbf{1} - 1).$$

As only the 2-dimensional case is considered explicitly, we can express the constraint $x^T \mathbf{1} = 1$ in the form $x = (z, 1-z)^T$. Incorporating this representation into the equation

$$\frac{\partial L}{\partial x} = -\hat{u} + \delta \frac{x}{\|x\|} + \lambda \mathbf{1} \stackrel{!}{=} 0$$

yields

$$-\begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \end{pmatrix} + \delta \frac{1}{\sqrt{z^2 + (1-z)^2}} \cdot \begin{pmatrix} z \\ 1-z \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} \stackrel{!}{=} 0,$$

i.e. the system of equations

$$-\hat{u}_1 + \frac{\delta z}{\sqrt{z^2 + (1-z)^2}} + \lambda \stackrel{!}{=} 0 \tag{1}$$

$$-\hat{u}_2 + \frac{\delta(1-z)}{\sqrt{z^2 + (1-z)^2}} + \lambda \stackrel{!}{=} 0. \tag{2}$$

From equation (2) we have

$$\lambda = \hat{u}_2 - \frac{\delta(1-z)}{\sqrt{z^2 + (1-z)^2}}$$

which changes equation (1) to

$$-\hat{u}_1 + \frac{\delta z}{\sqrt{z^2 + (1-z)^2}} + \hat{u}_2 - \frac{\delta(1-z)}{\sqrt{z^2 + (1-z)^2}} \stackrel{!}{=} 0$$

or equivalently

$$\frac{\delta(2z-1)}{\sqrt{z^2 + (1-z)^2}} = \hat{u}_1 - \hat{u}_2. \tag{3}$$

Now we distinguish two cases:

- $\hat{u}_1 = \hat{u}_2$.
  This results in the equation $2z - 1 = 0$, or $z = \frac{1}{2}$, i.e. the optimal solution in this case is the vector $x = (\frac{1}{2}, \frac{1}{2})^T$. The optimal objective value then is

$$\begin{aligned} f^*_{LRC}(\hat{u}, \delta) &= -x^T \hat{u} + \delta \|x\| \\ &= -\hat{u}_1 + \delta \sqrt{\frac{1}{2}}. \end{aligned}$$

- $\hat{u}_1 \neq \hat{u}_2$.
  Starting from Equation (3) we can do the following equivalent reformulations:

$$\begin{aligned} & \frac{\delta^2(4z^2 - 4z + 1)}{z^2 + (1-z)^2} = (\hat{u}_1 - \hat{u}_2)^2 \\ \Leftrightarrow \quad & \delta^2(4z^2 - 4z + 1) = 2z^2 \underbrace{(\hat{u}_1 - \hat{u}_2)^2}_{=:c} - 2z(\hat{u}_1 - \hat{u}_2)^2 + (\hat{u}_1 - \hat{u}_2)^2 \\ \Leftrightarrow \quad & z^2 \cdot 2(2\delta^2 - c) + z \cdot (-2)(2\delta^2 - c) + (\delta^2 - c) = 0 \end{aligned}$$

which thus give the possible solutions

$$\begin{aligned} z &= \frac{2(2\delta^2 - c) \pm \sqrt{4(2\delta^2 - c)^2 - 4 \cdot 2(2\delta^2 - c)(\delta^2 - c)}}{4(2\delta^2 - c)} \\ &= \frac{1}{2} \pm \frac{\sqrt{c(2\delta^2 - c)}}{2(2\delta^2 - c)} \\ &= \frac{1}{2} \pm \frac{\sqrt{c}}{2\sqrt{2\delta^2 - c}} \\ &= \frac{1}{2} \pm \frac{|\hat{u}_1 - \hat{u}_2|}{2\sqrt{2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2}} \end{aligned}$$

where a solution only exists if the discriminant is non-negative, i.e.

$$2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2 \geq 0$$

or

$$\delta \geq \frac{|\hat{u}_1 - \hat{u}_2|}{\sqrt{2}}.$$

Now that having a candidate solution, we finally need to determine the bounds on δ such that $0 \leq z \leq 1$ holds, i.e. that $x = (z, 1 - z)^T$ represents a feasible solution of the optimization problem. Because of the symmetry of z and $(1 - z)$, it suffices to investigate the case

$$\frac{1}{2} + \frac{|\hat{u}_1 - \hat{u}_2|}{2\sqrt{2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2)}} \leq 1$$

which can be simplified to

$$\begin{aligned} \Leftrightarrow \quad & \frac{|\hat{u}_1 - \hat{u}_2|}{2\sqrt{2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2)}} \leq \frac{1}{2} \\ \Leftrightarrow \quad & |\hat{u}_1 - \hat{u}_2| \leq \sqrt{2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2)} \\ \Rightarrow \quad & 2(\hat{u}_1 - \hat{u}_2)^2 \leq 2\delta^2 \\ \Rightarrow \quad & \delta \geq |\hat{u}_1 - \hat{u}_2|. \end{aligned}$$

(Note that in this case also the condition $\delta \geq \frac{|\hat{u}_1 - \hat{u}_2|}{\sqrt{2}}$ holds, i.e. the reformulations above are admissible.)

In the case where $\delta < |\hat{u}_1 - \hat{u}_2|$, the variable z reaches one of its bounds 0 or 1, and thus we again get one of the extreme solutions $(1, 0)^T$ or $(0, 1)^T$.

Summarizing the previous calculations, we have the following result for the

optimal solution of the local robust counterpart program:

$$\mathcal{F}^*_{LRC}(\hat{u},\delta) =$$
$$= \begin{cases} \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \text{if } \hat{u}_1 \geq \hat{u}_2 + \delta \\ \begin{pmatrix} \min\left\{\frac{1}{2} + \frac{|\hat{u}_1-\hat{u}_2|}{2\sqrt{2\delta^2-(\hat{u}_1-\hat{u}_2)^2}}; 1\right\} \\ \max\left\{\frac{1}{2} - \frac{|\hat{u}_1-\hat{u}_2|}{2\sqrt{2\delta^2-(\hat{u}_1-\hat{u}_2)^2}}; 0\right\} \end{pmatrix} & \text{if } \hat{u}_2 + \delta > \hat{u}_1 > \hat{u}_2 \\ \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} & \text{if } \hat{u}_1 = \hat{u}_2 \\ \begin{pmatrix} \max\left\{\frac{1}{2} - \frac{|\hat{u}_1-\hat{u}_2|}{2\sqrt{2\delta^2-(\hat{u}_1-\hat{u}_2)^2}}; 0\right\} \\ \min\left\{\frac{1}{2} + \frac{|\hat{u}_1-\hat{u}_2|}{2\sqrt{2\delta^2-(\hat{u}_1-\hat{u}_2)^2}}; 1\right\} \end{pmatrix} & \text{if } \hat{u}_1 < \hat{u}_2 < \hat{u}_1 + \delta \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \text{if } \hat{u}_1 + \delta \leq \hat{u}_2. \end{cases}$$

Before calculating the corresponding optimal objective value, note that

$$\begin{aligned} \|x\|^2 &= z^2 + (1-z)^2 \\ &= \frac{1}{4} \pm \frac{|\hat{u}_1 - \hat{u}_2|}{2\sqrt{2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2}} + \frac{(\hat{u}_1 - \hat{u}_2)^2}{4(2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2)} \\ &\quad + \frac{1}{4} \mp \frac{|\hat{u}_1 - \hat{u}_2|}{2\sqrt{2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2}} + \frac{(\hat{u}_1 - \hat{u}_2)^2}{4(2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2)} \\ &= \frac{1}{2} + \frac{(\hat{u}_1 - \hat{u}_2)^2}{2(2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2)} \\ &= \frac{\delta^2}{2\delta^2 - (\hat{u}_1 - \hat{u}_2)^2} \end{aligned}$$

In the case where $\hat{u}_1 > \hat{u}_2$ and $\delta \geq |\hat{u}_1 - \hat{u}_2|$ the optimal solution x is given by

$$x = \begin{pmatrix} \frac{1}{2} + \frac{|\hat{u}_1-\hat{u}_2|}{2\sqrt{2\delta^2-(\hat{u}_1-\hat{u}_2)^2}} \\ \frac{1}{2} - \frac{|\hat{u}_1-\hat{u}_2|}{2\sqrt{2\delta^2-(\hat{u}_1-\hat{u}_2)^2}} \end{pmatrix}$$

and thus the objective value calculates as

$$\begin{aligned} f^*_{LRC}(\hat{u},\delta) = & -\frac{1}{2}(\hat{u}_1+\hat{u}_2) - \frac{|\hat{u}_1-\hat{u}_2|}{2\sqrt{2\delta^2-(\hat{u}_1-\hat{u}_2)^2}}(\hat{u}_1-\hat{u}_2) \\ & + \delta\sqrt{\frac{\delta^2}{2\delta^2-(\hat{u}_1-\hat{u}_2)^2}} \\ = & -\frac{1}{2}(\hat{u}_1+\hat{u}_2) - \frac{1}{2}\frac{(\hat{u}_1-\hat{u}_2)^2}{\sqrt{2\delta^2-(\hat{u}_1-\hat{u}_2)^2)}} \\ & + \frac{\delta^2}{\sqrt{2\delta^2-(\hat{u}_1-\hat{u}_2)^2}} \\ = & -\frac{1}{2}(\hat{u}_1+\hat{u}_2) + \frac{2\delta^2-(\hat{u}_1-\hat{u}_2)^2}{2\sqrt{2\delta^2-(\hat{u}_1-\hat{u}_2)^2}} \\ = & -\frac{1}{2}(\hat{u}_1+\hat{u}_2) + \frac{1}{2}\sqrt{2\delta^2-(\hat{u}_1-\hat{u}_2)^2}. \end{aligned}$$

The case where $\hat{u}_1 < \hat{u}_2$ is done analogously, and the cases where it holds that $\delta < |\hat{u}_1 - \hat{u}_2|$ result in extreme optimal solutions, i.e. in an optimal objective value of

$$-\hat{u}_1+\delta \quad \text{or, respectively,} \quad -\hat{u}_2+\delta.$$

Bibliography

[1] Anderson, T. W., An Introduction to Multivariate Statistical Analysis, *John Wiley & Sons*, 1958.

[2] Artzner, P. and Delbaen, F. and Eber, J.-M. and Heath, D., Coherent measures of risk. *Mathematical Finance*, **9**, **3**, 1999.

[3] Bank, B. et al., Non-Linear Parametric Optimization. *Birkhäuser, Basel*, 1983.

[4] Ben-Tal, A. and Nemirovski, A., Robust convex optimization. *Mathematics of Operations Research*, **23**, **4**, 1998.

[5] Ben-Tal, A. and Nemirovski, A., Robust Solutions of Uncertain Linear Programs. *OR letters*, **25**, **1**, 1999.

[6] Ben-Tal, A. and Margalit, T. and Nemirovski, A., Robust Modeling of Multi-Stage Portfolio Problems. In: H. Frenk, K. Roos, T. Terlaky, S. Zhang (Eds.), High Performance Optimization, *Kluwer Academic Publishers*, 2000.

[7] Ben-Tal, A. and Nemirovski, A., Robust solutions of Linear Programming problems contaminated with uncertain data, *Mathematical Programming*, **88**, **3**, 2000.

[8] Ben-Tal, A. and Nemirovski. A., Robust Optimization - Methodology and Applications. *Mathematical Programming*, **92**, **3**, 2002.

[9] Ben-Tal, A., Roos, C. and Nemirovski. A., Robust solutions to uncertain quadratic and conic-quadratic problems, *SIAM Journal on Optimization*, **13**, **2**, 2002.

[10] Ben-Tal, A. and Boyd, S. and Nemirovski, A,. Extending Scope of Robust Optimization: Comprehensive Robust Counterpart of Uncertain Problems, *Mathematical Programming*, **107**, **1/2**, 2006.

[11] Ben-Tal, A. and Nemirovski, A,. Selected topics in robust convex optimization, To appear in: *Mathematical Programming*.

[12] Bertsimas, D. and Sim, M., Tractable Approximations to Robust Conic Optimization Problems, *Mathematical Programming*, **107**, **1/2**, 2006.

[13] Best, M. J. and Grauer, R. R., On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results, *The Review of Financial Studies*, **4**, **2**, 1991.

[14] Bliman, P.-A. and Prieur, C., On existence of smooth solutions of parameter-dependent convex programming problems, *Proceedings of 16th International Symposium on Mathematical Theory of Networks and Systems (MTNS2004)*, 2004.

[15] Black, F. and Litterman, R., Global Portfolio Optimization, *Financial Analysts Journal*, **48**, **5**, 1992.

[16] Bonnans, J. F. and Shapiro, A., Optimization Problems with perturbations, A guided tour. *Rapport de recherche n° 2872, INRIA Rocquencourt*, 1996.

[17] Bonnans, J. F. and Cominetti, R. and Shapiro, A., Sensitivity analysis of optimization problems under second order regular constraints. *Rapport de recherche n° 2989, INRIA Rocquencourt*, 1996.

[18] Bonnans, J. F. and Shapiro, A., Perturbation Analysis of Optimization Problems. *Springer, New York*, 2000.

[19] Boyd, S. and Vandenberghe, L., Convex Optimization. *Cambridge University Press*, 2004.

[20] Ceria, S. and Stubbs, R. A., Incorporating estimation errors into portfolio selection: Robust portfolio construction, *Journal of Asset Management*, **7**, **2**, 2006.

[21] Chopra, V. K. and Ziemba, W. T., The Effect of Errors in Means, Variances and Covariances on Optimal Portfolio Choice, *Journal of Portfolio Management*, **19**, **2** 1993.

[22] Cornuejols, G. and Tütüncü, R. H. Optimization Methods in Finance, *Cambridge University Press*, 2006.

[23] DeMiguel, V. and Nogales, F., Portfolio Selection with Robust Estimates of Risk, *SSRN working paper*, 2006 [online]. Available from: http://papers.ssrn.com

[24] Dontchev, A. L., Technical Note: A Proof of the Necessity of Linear Independence Condition and Strong Second-Order Sufficient Optimality Condition for Lipschitzian Stability in Nonlinear Programming. *Journal of Optimization Theory and Applications*, **98**, **2**, 1998.

[25] Efron, B. and Morris, C., Data Analysis Using Stein's Estimator and Its Generalizations. *Journal of the American Statistical Association*, **70**, 1975.

[26] El-Ghaoui, L. and Oustry, F. and Lebret, H., Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, **9**, **1**, 1998.

[27] El-Ghaoui, L. and Oks, M. and Oustry, F., Worst-Case Value-at-Risk and Robust Portfolio Optimization: A Conic Programming Approach. *Operations Research*, **51**, **4**, 2003.

[28] Fabian, M. et al., Functional analysis and infinite-dimensional geometry. *Springer, New York*, 2001.

[29] Fang, K.-T. and Kotz, S. and Ng, K.-W., Symmetric Multivariate and Related Distributions, *Chapman and Hall, London*, 1990.

[30] Fang, K.-T. and Zhang, Y.-T., Generalized multivariate analysis, *Science Press, Beijing*, 1990.

[31] Frahm, G., Generalized Elliptical Distributions: Theory and Applications, *Dissertation Thesis, Köln*, 2004.

[32] Goberna, M. A. and López, M. A., Linear Semi-infinite Optimization, *John Wiley & Sons, Chichester*, 1998.

[33] Goldfarb, D. and Iyengar, G., Robust Portfolio Selection Problems, *Mathematics of Operations Research* **28**, **1**, 2003.

[34] Goodall, C., M-Estimators of Location: An Outline of the Theory, In: D. C. Hoaglin, F. Mosteller, J. W. Tukey (Eds.), Understanding Robust and Exploratory Data Analysis, *John Wiley & Sons, New York*, 1983.

[35] Gugat, M., Parametrische Optimierung, Vorlesungsskript, 2003. Available from: http://www2.am.uni-erlangen.de/∼gugat.

[36] Hadamard, J., Lectures on Cauchy's Problem in Linear Partial Differential Equations, *Yale University Press, New Haven*, 1923.

[37] Huber, P. J., Robust Estimation of a Location Parameter, *The Annals of Mathematical Statistics*, **35**, **1**, 1964.

[38] Huber, P. J., Robust Statistics, *John Wiley & Sons, New York*, 1981.

[39] Hult, H. and Lindskog, F., Multivariate extremes, aggregation and dependence in elliptical distributions, *Advances in Applied Probability*, **34**, **3**, 2002.

[40] Ingersoll, J. E., Theory of Financial Decision Making, *Rowman & Littlefield, Maryland*, 1987.

[41] Jacod, J. and Protter, P., Probability Essentials, Second Edition, *Springer, Berlin*, 2003.

[42] Jahn, J., Introduction to the Theory of Nonlinear Optimization, *Springer, Berlin*, 1994.

[43] Jänich, K., Topologie, Vierte Auflage, *Springer, Berlin*, 1994.

[44] Jobson, J. D. and Korkie, B., Estimation for Markowitz efficient portfolios, *J. Amer. Statist. Assoc.*, **75**, **371**, 1980.

[45] Jobson, J. D. and Korkie, B., Putting Markowitz theory to work, *Journal of Portfolio Management*, **7**, **4**, 1981.

[46] Johnson, N. L. and Kotz, S. and Balakrishnan, N., Continuous Univariate Distributions, Volume 1, Second Edition, *John Wiley & Sons, New York*, 1994.

[47] Jorion, P., Bayes-Stein Estimation for Portfolio Analysis, *Journal of Financial and Quantitative Analysis*, **21**, **3**, 1986.

[48] Jorion, P., Portfolio Optimization in Practice, *Financial Analysts Journal*, **48**, **1**, 1992.

[49] Kirsch, A., An Introduction to the Mathematical Theory of Inverse Problems, *Springer, New York*, 1996.

[50] Lauprête, G. J., Portfolio Risk Minimization under Departures from Normality. *PhD Thesis, MIT*, 2001.

[51] Lauprête, G. J. and Samarov, A. M. and Welsch, R. E., Robust Portfolio Optimization, *Metrika*, **55**, **2**, 2002.

[52] Lobo, M. S., Robust and Convex Optimization with Applications in Finance, *PhD Thesis, Stanford University*, 2000.

[53] Luenberger, D. G., Optimization by vector space methods. *John Wiley & Sons*, 1969.

[54] Lutgens, F. and Sturm, J., Robust one period option modelling, *Discussion Paper 114, Tilburg University, Center for Economic Research*, 2002.

[55] Lutgens, F., Robust Portfolio Optimization, *PhD Thesis, Maastricht University*, 2004.

[56] Markowitz, H., Portfolio Selection, *Journal of Finance*, **7**, **1**, 1952.

[57] Meucci, A., Risk and Asset Allocation, *Springer, Berlin*, 2005.

[58] Meucci, A., Robust Bayesian Allocation, *SSRN working paper*, 2005 [online]. Available from: http://papers.ssrn.com.

[59] Meyer, C. D., Matrix Analysis and Applied Linear Algebra, *SIAM, Philadelphia*, 2000.

[60] Michaud, R. O., Efficient Asset Management, *Harvard Business School Press, Boston*, 1998.

[61] Middelkamp, C., Untersuchung eines Verfahrens zur robusten Portfoliooptimierung unter elliptischen Verteilungsannahmen, *Diplomarbeit, Technische Universität München*, 2007.

[62] Mori, H., Finite sample properties of estimators for the optimal portfolio weight, *Journal of the Japan Statistical Society*, **34**, **1**, 2004.

[63] Natarajan, K. and Pachamanova, D. and Sim, M., Constructing Risk Measures from Uncertainty Sets. *Working Paper*, 2005. Available from: http://www.optimization-online.org/DB_HTML/2005/07/1184.html.

[64] Perret-Gentil, C. and Victoria-Feser, M.-P., Robust Mean-Variance Portfolio Selection, *FAME Research Paper no 140*, 2004.

[65] Press, S. J., Applied Multivariate Analysis, *Holt, Rinehart and Winston, Inc., New York*, 1972.

[66] Qian, E. and Gorman, S., Conditional Distribution in Portfolio Theory, *Financial Analysts Journal*, **57**, **2**, 2001.

[67] Reemtsen, R. and Rückmann, J.-J. (eds.), Semi-infinite Programming, *Kluwer Academic Publishers, Dordrecht*, 1998.

[68] Repovš, D. and Semenov, P.V., Continuous Selections of Multivalued Mappings, *Kluwer Academic Publishers, Dordrecht*, 1998.

[69] Rinne, H., Taschenbuch der Statistik, 3., vollständig überarbeitete und erweiterte Auflage, *Harri Deutsch GmbH, Frankfurt*, 2003.

[70] Rockafellar, R. T., Convex Analysis, *Princeton University Press*, 1997.

[71] Rockafellar, T. R. and Uryasev, S. and Zabarankin, M., Deviation Measures in Risk Analysis and Optimization, *University of Florida, Department of Industrial & Systems Engineering Working Paper No. 2002-7*, 2002. Available from: http://ssrn.com/abstract=365640.

[72] Rockafellar, R. T. and Uryasev, S., Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance*, **26**, 2002.

[73] Rockafellar, R. T. and Uryasev, S. and Zabarankin, M., Generalized Deviations in Risk Analysis, *Finance & Stochastics*, **10**, **1**, 2006.

[74] Roman, S., Advanced Linear Algebra, *Springer, New York*, 1992.

[75] Scherer, B., Portfolio Resampling: Review and Critique, *Financial Analysts Journal*, **58**, **6**, 2002.

[76] Scherer, B., Portfolio Construction and Risk Budgeting, 2nd edition, *Risk Waters Group Ltd., London*, 2004.

[77] Schöttle, K. and Werner, R., Towards reliable efficient frontiers, *Journal of Asset Management*, **7**, **2**, 2006.

[78] Shorack, G. R. and Wellner, J. A., Empirical Processes with Applications to Statistics, *John Wiley & Sons, New York*, 1986.

[79] Tütüncü, R. H. and Koenig, M., Robust Asset Allocation, *Annals of Operations Research*, **132**, **1-4**, 2004.

[80] van der Vaart, A. W., Asymptotic Statistics, *Cambridge University Press*, 1998.

[81] Victoria-Feser, M.-P., Robust Portfolio Selection, *Paper Series, 2000.14, Ecole des Hautes Etudes Commerciales, Universite de Geneve*, 2000.

List of Figures

Notation

X	General constraint set for x
$\mathcal{U} \subset \mathbb{R}^d$	General uncertainty set
$\mathcal{U}_\delta(\hat{u})$	Uncertainty set centered at $\hat{u}$ and of size δ
f^*	Optimal value function or extreme value function of the original optimization problem
f^*_{LRC}	Optimal value function of the local robust counterpart problem
$\mathcal{F}$	Feasibility set of the original problem
$\mathcal{F}_{LRC}$	Feasibility set of the local robust counterpart problem
$\mathcal{F}^S$	Set of Slater points of the original problem
$\mathcal{F}^S_{LRC}$	Set of Slater points of the local robust counterpart problem
$\mathcal{F}^*$	Set of optimal solutions of the original problem
$\mathcal{F}^*_{LRC}$	Set of optimal solutions of the local robust counterpart problem
$\mathcal{F}^*_\varepsilon$	Set of ε-optimal solutions of the original problem
ζ	Selection function
H-lsc	Abbreviation for: Hausdorff lower semicontinuous
H-usc	Abbreviation for: Hausdorff upper semicontinuous
B-lsc	Abbreviation for: Berge lower semicontinuous
B-usc	Abbreviation for: Berge upper semicontinuous
Γ	A general multi-valued mapping
$\mathcal{P}(X)$	Power set of X
$V_\varepsilon(u), V_\varepsilon(S)$	ε-neighborhood of a point u or a set S
$o(\alpha^k)$	o-notation, defined as $f(\alpha) \in o(\alpha^k) \quad \Leftrightarrow \quad \lim_{\alpha \to 0} \frac{f(\alpha)}{\alpha^k} = 0.$
ϕ	Characteristic generator of a spherical or elliptical distribution
ψ	Characteristic function
ξ	Density generator of a spherical or elliptical distribution
φ	Density function of a (continuous) random variable

$x \in \mathbb{R}^n$	In applications: vector of portfolio weights
x^*_{cl}	Optimal solution to the classical portfolio optimization problem
x^*_{rob}	Optimal solution to the robust portfolio optimization problem
$\mu \in \mathbb{R}^n$	Vector of expected returns of the assets
$\Sigma \in \mathbb{R}^{n \times n}$	Covariance matrix of the asset returns
$\hat{\mu}, \hat{\Sigma}$	Estimators for μ and Σ, respectively
$\mathbf{1}$	Vector of ones in the appropriate dimension

cl	Closure of a set
pdf	Abbreviation for: probability density function
cdf	Abbreviation for: cumulated density function
MVP	Abbreviation for: minimum variance portfolio
MRP	Abbreviation for: maximum return portfolio

$Z_n \xrightarrow{\mathbf{P}} Z$	Z_n converges in probability to Z
$Z_n \xrightarrow{\text{a.s.}} Z$	Z_n converges almost surely to Z
$Z_n \xrightarrow{d} Z$	Z_n converges in distribution to Z