

Consistency of robust optimization with application to portfolio optimization

Ralf Werner

Hochschule München, Fakultät für Informatik und Mathematik, Lothstr. 64, 80335 München
email: werner@cs.hm.edu

In recent years the robust counterpart approach, introduced and made popular by Ben-Tal, Nemirovski and El Ghaoui, gained more and more interest among both academics and practitioners. However, to the best of our knowledge, only very few results on the relationship between the original problem instance and the robust counterpart have been established. This exposition aims at closing this gap by showing that the robust counterpart to an already well-posed problem remains well-posed under some mild regularity and uniqueness assumption on the solution of the original problem instance. As a consequence, sufficient conditions will be established under which the solution of the robust counterpart converges to the original solution, if the level of robustification is decreased to zero. Based on the well-posedness of the robust counterpart, it will also be demonstrated how any consistent plug-in estimator can be supplemented by a corresponding consistent robust estimator based on a proper choice of the confidence set of the plug-in estimator. Finally, this consistency result leads to a generalization of already known consistency results in the framework of mean-variance portfolio optimization.

Key words: robust optimization; portfolio optimization; consistency

MSC2000 Subject Classification: Primary: 90C25, 90C90 ; Secondary: 90C31, 62P05, 62H12

OR/MS subject classification: Primary: programming, nonlinear ; Secondary: statistics, estimation

1. Motivation and introduction. Most optimization problems arising in practice are in fact parametric ones. These problems usually depend on some data entering both the objective and the constraints of the optimization problem. In this sense, the optimal solution is then depending on the given parameters. In most cases the data is not known exactly, but uncertain, either due to measurement errors or rounding procedures, or because these values even stem from estimation procedures. In most practical situations a – preferably unique – solution which is obtained for a specific parameter needs to be implemented, e.g. an optimal mechanical design is constructed or an optimal portfolio is set up. Hence it is desired that this implemented solution is close to the optimal solution if the obtained data is sufficiently close to the true data. In other words, practitioners favor *well-posed* problems in the sense of Hadamard, see Hadamard [15]. Below, we therefore focus on *well-posed problems* meaning that the following three requirements are met (see Theorem 1.2 for a more exact definition):

- (i) *existence* of an optimal solution,
- (ii) *uniqueness* of the optimal solution,
- (iii) *stability* of the optimal solution.

In this context, stability of the optimal solution simply means continuity of the optimal solution with respect to the parameter, cf. Section 1.7. This can obviously only be the case if (i) and (ii) hold. A lot of research has been dedicated to find necessary and sufficient conditions for (i) to (iii), see e.g. Bank et al [1] and Bonnans and Shapiro [6, 7] for an overview. However, even for well-posed problems the optimal solution might become infeasible under slight perturbations of the initial parameter. For this reason, Ben-Tal, El Ghaoui and Nemirovski introduced the concept of the robust counterpart which guarantees feasibility under reasonable small perturbations of the original parameter, see Ben-Tal, El Ghaoui and Nemirovski [2] for an overview of the theory of robust optimization.

Although there exists extensive literature on robust optimization, see the references in [2], there is almost no result which links the robust counterpart to its original problem. To the best of our knowledge, the only exceptions are the very early results by Ben-Tal and Nemirovski in [3]. Therein, Theorem 2.1 relates the optimal value of the robust counterpart to the worst of all instances in case of affine uncertainty. For uncertain LPs under concave uncertainty, Ben-Tal and Nemirovski have shown in Section 2.3 in [3] that under very mild conditions the optimal value of the robust counterpart converges to the optimal value of the original LP if the uncertainty set shrinks to the original data point. Although no explicit statement about the convergence of the optimal solution of the robust counterpart to a solution of the nominal instance for decreasing level of robustness (i.e. decreasing diameter of the uncertainty set) was made, it can be observed from the proof of Proposition 2.3 in [3] that any cluster point of the optimal robust solutions is optimal in the original LP.

Subsequently, this exposition will try to close the gap concerning the relation between the robust counterpart and the nominal instances. In particular, we will make the following contributions to the topic of robust optimization:

- It will be shown that the robust counterpart has the same structure as the original problem, i.e. the robust counterpart remains a convex parametric program like the original problem instance, however, the uncertain parameter is replaced by an uncertainty set.
- The robust counterpart is again a well-posed problem if the original problem is already well-posed and if some additional mild conditions are satisfied. In this case, it is further shown that the solutions of the robust counterpart converge to a solution of the original problem, if the uncertainty set is reduced to the original parameter.
- In addition, a counterexample will be provided which demonstrates that solely well-posedness of the original problem is not sufficient for well-posedness of the robust counterpart and therefore, additional requirements have to be satisfied.

Focusing on the statistical framework of point estimators, a supplementary robustified plug-in estimator will be established in addition to the usual plug-in estimator based on the robust counterpart. The previous continuity results will allow to prove consistency of the robustified estimator, if the original estimator has already been consistent. Considering the particular application of mean-variance portfolio optimization, it will be easily possible to establish consistency of a variety of portfolio estimators. To the best of our knowledge, this extends known consistency results of mean-variance optimization to very general portfolio and risk constraints. Further, consistency will also carry over from the standard mean-variance problem to its robust counterpart.

The rest of the paper is organized as follows: in the remainder of the first section, the main results on continuity of optimal solutions of convex optimization problems with respect to some parameter will be recalled. In Section 2, after a succinct introduction of the robust counterpart, it will be shown that the parametric structure of the robust problem resembles the original problem based on a suitable reformulation. This will allow to carry over the results from Section 1, especially well-posedness of the original problem, to the robust counterpart under very mild assumptions. Two carefully constructed examples will demonstrate that these mild assumptions can hardly be relaxed any further. The well-posedness will then be embedded into a statistical framework in Section 3, where consistency of plug-in estimators together with consistency of robustified plug-in estimators is derived. An application of these results to mean-variance optimization will allow to generalize known consistency results in Section 4, before an outlook to possible directions of future research is given.

1.1 Setup. Throughout the paper we will consider the following convex finite dimensional conic optimization problem

$$\begin{aligned} \min_{x \in X} \quad & f(x, u) \\ \text{s.t.} \quad & g(x, u) \in -K, \end{aligned} \tag{P_u}$$

which covers a large variety of convex optimization problems, like LPs, SOCPs or SDPs. For the purpose of this exposition, a few mild assumptions on (P_u) are necessary, which will be introduced in more detail in Sections 1.2, Sections 1.4 and Sections 1.6. We will see later on that these assumptions are for example satisfied for the mean-variance portfolio optimization framework considered in Section 4.

1.2 Basic assumptions. Let us assume throughout that the set $X \subset \mathbb{R}^n$ is non-empty, convex and compact. Further, let $u \in \bar{U} \subset \mathbb{R}^d$ denote the uncertain data within a *global uncertainty set* \bar{U} which is also assumed to be non-empty, convex¹ and compact. Let $K \subset \mathbb{R}^m$ be a proper ordering cone with $\text{int}(K) \neq \emptyset$ i.e. we only consider parameter dependence for the inequality constraints. In addition, let the objective $f : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}$ and the constraint $g : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}^m$ be continuous in (x, u) . Finally, to obtain a convex conic problem (P_u) let us assume that f is convex in x and g is K -convex in x for all $u \in \bar{U}$. Please note that we do not impose any structure on the way the parameter enters the objective or constraint function.

¹As noted in Ben-Tal et al [2], \bar{U} can be replaced by $\text{conv}(U)$ if g is K -convex in u for all x .

1.3 Basic properties. From these basic assumptions, it already follows that there exists at least one optimal solution $x^*(u) \in X$ for each $u \in \bar{U}$. Let us briefly go through the main known results concerning stability of the feasible set and especially the set of optimal solution, before these are generalized to the robust counterpart in Section 2.

DEFINITION 1.1 *Let $\mathcal{K}(\mathbb{R}^n)$ be the space of all non-empty, convex and compact subsets of \mathbb{R}^n . Further, let us introduce*

- the feasible set mapping

$$\Phi : \bar{U} \rightarrow \mathcal{K}(\mathbb{R}^n), u \mapsto \Phi(u) := \{x \in X \mid g(x, u) \in -K\},$$

- the optimal value function

$$f^* : \bar{U} \rightarrow \mathbb{R}, u \mapsto f^*(u) := \min_{x \in \Phi(u)} f(x, u), \quad \text{and}$$

- the optimal set mapping

$$\mathcal{S} : \bar{U} \rightarrow \mathcal{K}(\mathbb{R}^n), u \mapsto \mathcal{S}(u) := \{x \in \Phi(u) \mid f(x, u) \leq f^*(u)\}.$$

To study the continuity properties of these mappings, let us recall the notion of Hausdorff distance and Hausdorff continuity:

DEFINITION 1.2 *Let $A, B \in \mathcal{K}(\mathbb{R}^n)$. Then the Hausdorff excess of A over B is defined as $e_H(A, B) := \sup_{a \in A} d(a, B)$ where the distance of the point a to the set B is given as $d(a, B) := \inf_{b \in B} \|a - b\|$ for some norm on \mathbb{R}^n .*

It is known that due to the missing symmetry, i.e. due to $e_H(A, B) \neq e_H(B, A)$, the excess does not define a metric on the space $\mathcal{K}(\mathbb{R}^n)$.

DEFINITION 1.3 *Let $A, B \in \mathcal{K}(\mathbb{R}^n)$. Then the Hausdorff distance between A and B , given by $d_H(A, B) := \max(e_H(A, B), e_H(B, A))$, induces a metric on the space $\mathcal{K}(\mathbb{R}^n)$.*

Based on the Hausdorff excess and the Hausdorff distance, Hausdorff upper and lower semicontinuity of set-valued mappings can be characterized. It is easy to see that the following characterization coincides with the definitions in Bank et al [1], p. 25.

DEFINITION 1.4 *A sequence $(A_k)_{k \in \mathbb{N}}$ with $A_k \in \mathcal{K}(\mathbb{R}^n)$ is called Hausdorff convergent to $A \in \mathcal{K}(\mathbb{R}^n)$, in short $A_k \xrightarrow{H} A$, if $\lim_{k \rightarrow \infty} d_H(A_k, A) = 0$. In the same manner a set-valued mapping $\mathcal{F} : \mathbb{R}^d \rightarrow \mathcal{K}(\mathbb{R}^n)$ is called*

- Hausdorff upper semicontinuous in x if $e_H(\mathcal{F}(x_k), \mathcal{F}(x)) \rightarrow 0$ for all sequences $x_k \rightarrow x$,
- Hausdorff lower semicontinuous in x if $e_H(\mathcal{F}(x), \mathcal{F}(x_k)) \rightarrow 0$ for all sequences $x_k \rightarrow x$,
- Hausdorff continuous in x if $\mathcal{F}(x_k) \xrightarrow{H} \mathcal{F}(x)$ for all sequences $x_k \rightarrow x$, i.e. \mathcal{F} is Hausdorff continuous if it is both Hausdorff upper and lower semicontinuous.

Please note that in this framework, Hausdorff semicontinuity (sometimes also called hemicontinuity) is equivalent to Berge semicontinuity in the sense of Bank et al [1], p. 25. This follows from the compactness of X and Bank et al [1], Lemma 2.2.3, and therefore, these notions of semicontinuity can subsequently be used synonymously. Let us summarize the main continuity results in this context in the following theorem.

THEOREM 1.1 *Under the basic assumptions of Section 1.2 it holds that*

- (i) the feasible set mapping Φ is closed and Hausdorff upper semicontinuous on \bar{U} ,
- (ii) the optimal value function f^* is lower semicontinuous on \bar{U} , and

- (iii) the optimal set mapping \mathcal{S} is Hausdorff upper semicontinuous on \bar{U} if and only if the optimal value function f^* is upper semicontinuous on \bar{U} .

PROOF. For the closedness and Hausdorff upper semicontinuity of the feasible set mapping Φ see Bank et al [1], Theorem 3.1.1 and Theorem 3.1.2. Lower semicontinuity of the optimal value function f^* follows from Bank et al [1], Theorem 4.2.1(1). The relation between the optimal set mapping \mathcal{S} and the optimal value function f^* is again due to Bank et al [1], Corollary 4.2.1.1. \square

1.4 Uniqueness assumption. As pointed out by Kirsch [21], p. 10, problems without a unique solution usually miss some crucial additional information. Obviously, in such a situation it does not make sense to consider well-posedness, i.e. stability of the optimal solution with respect to variations of the parameter u , as the solution is not unique. Therefore the following uniqueness assumption is usually imposed:

$$\text{The optimal solution set } \mathcal{S}(u) \text{ is a singleton } \mathcal{S}(u) = \{x^*(u)\} \text{ for all } u \in \bar{U}. \quad (SA)$$

Assumption (SA) is a common assumption in the investigation of well-posedness, see for example Bonnans and Shapiro [7], but as can be observed in Example 2.1, it is too weak for the purpose of this exposition. To guarantee uniqueness of the optimal solution $x^*(u)$, i.e. to guarantee that $\mathcal{S}(u)$ is a singleton, it could also be assumed that f is *strictly quasiconvex* in x for all $u \in \bar{U}$:

$$\forall u \in \bar{U}, \forall 0 < \lambda < 1, \forall x \neq y: \quad f(\lambda x + (1 - \lambda)y, u) < \max(f(x, u), f(y, u)). \quad (CA)$$

Obviously, if f is strictly convex in x , then it is also strictly quasiconvex. Although strict quasiconvexity might seem to be a rather strong assumption, we will see in the following that it is for example fulfilled for the particular application of portfolio optimization which we will consider later.

REMARK 1.1 *If no uniqueness assumption can be reasonably assumed, as last resort, problem (P_u) could still be Tikhonov-regularized, i.e. $f(x, u)$ could be replaced by*

$$f_\gamma(x, u) := f(x, u) + \gamma \|x - x_0\|_2^2$$

for some $x_0 \in \mathbb{R}^n$. It is obvious that f_γ is strictly convex and that therefore the regularized problem $(P_{u,\gamma})$ possesses a unique solution $x_\gamma^(u)$ for all $\gamma > 0$. It is further easy to see that Hausdorff upper semicontinuity holds,*

$$e_H(\{x_\gamma^*(u)\}, \mathcal{S}(u)) = d(x_\gamma^*(u), \mathcal{S}(u)) \rightarrow 0 \text{ for } \gamma \rightarrow 0$$

i.e. each cluster point of the regularized solutions is a solution of the original problem, cf. Bonnans and Shapiro [7], Proposition 4.6. Under suitable second order growth conditions and appropriate smoothness conditions, statements concerning convergence rates can be obtained, where convergence rates of at least $\mathcal{O}(\sqrt{\gamma})$ can be achieved (see for example Bonnans and Shapiro [7], Section 4.4). Thus, instead of starting with the original problem (P_u) , the regularized problem $(P_{u,\gamma})$ for some small but fixed $\gamma > 0$ could be used as starting point for the following considerations alike.

1.5 Further properties. Based on the above uniqueness assumption (SA), the following proposition can easily be derived. Note that this means that Hausdorff lower semicontinuity of a set-valued map already follows from Hausdorff upper semicontinuity in this special case.

PROPOSITION 1.1 *Let the assumptions from Section 1.2 be satisfied and let $\mathcal{S}(u)$ be a singleton for all $u \in \bar{U}$, i.e. let (SA) hold. Then \mathcal{S} is Hausdorff continuous if one of the following equivalent conditions is fulfilled:*

- (i) \mathcal{S} is Hausdorff upper semicontinuous on \bar{U} ,
- (ii) f^* is upper semicontinuous (and thus continuous) on \bar{U} .

In other words, if (i) or (ii) is satisfied, then the mapping $u \mapsto x^(u)$ is continuous on \bar{U} .*

PROOF. The equivalence of (i) and (ii) is already known from Theorem 1.1(iii). Therefore, let us assume that \mathcal{S} is Hausdorff upper semicontinuous on \bar{U} . Using $\mathcal{S}(u) = \{x^*(u)\}$ this means that for all sequences $u_n \rightarrow u$ it holds that

$$\|x^*(u_n) - x^*(u)\| = e_H(\{x^*(u_n)\}, \{x^*(u)\}) = e_H(\mathcal{S}(u_n), \mathcal{S}(u)) \rightarrow 0,$$

which is exactly the usual continuity definition. \square

From this proposition we immediately see that some constraint qualification needs to be satisfied to guarantee continuity of the optimal value function: it is well-known that continuity of the optimal value function is closely related to the calmness of (P_u) , which is again closely related to a regularity condition (see e.g. Bonnans and Shapiro [7], pp. 99). For example, according to Bank et al [1], Corollary 4.4.1.3 or Bonnans and Shapiro [7], Section 2.5.4 we know that even in the simplest case $K = \mathbb{R}_+$, $f(x, u) = f(x)$ and $g(x, u) = g(x) - u$, continuity of f^* in $u = 0$ is equivalent to the existence of a Slater point for $u = 0$.

1.6 Regularity assumption. Motivated by the above considerations, let us assume that the following Slater condition holds:

$$\text{For } u \in \bar{U} \text{ there exists a Slater point } x_{Sl}(u) \in X \text{ such that } g(x_{Sl}(u), u) \in \text{int}(-K). \quad (SC_u)$$

1.7 Well-posedness of (P_u) . If the above Slater condition is satisfied, the following theorem gives a more precise definition of well-posedness than the previous colloquial description.

THEOREM 1.2 *Let the assumptions of Section 1.2 be fulfilled and let (SA) be satisfied in a neighborhood of a given $u \in \bar{U}$. If the Slater condition (SC_u) is satisfied in $u \in \bar{U}$, then*

- (i) *the feasible set mapping Φ is Hausdorff continuous,*
- (ii) *the optimal value function f^* is continuous, and*
- (iii) *the optimal set mapping \mathcal{S} is Hausdorff continuous*

in a neighborhood of u . This means that under (SC_u) problem (P_u) is well-posed, i.e. the set of optimal solutions is non-empty and contains exactly one element, i.e. $\mathcal{S}(u) = \{x^(u)\}$ and the mapping $u \mapsto x^*(u)$ is well-defined and continuous in some neighborhood of u .*

PROOF. First of all we note that existence of $x^*(u)$ follows from the assumptions in Section 1.2, and that uniqueness is explicitly assumed. Theorem 3.1.6 in Bank et al [1] yields that the feasible set mapping Φ is Hausdorff lower semicontinuous, and thus Hausdorff continuous in u . From this it follows in turn that f^* is upper semicontinuous, thus continuous, in u due to Bank et al [1], Theorem 4.2.2. With the above Proposition 1.1, the optimal set mapping \mathcal{S} and thus the optimal solution x^* is continuous in u . \square

2. Robust optimization. In the above setting, it is assumed that X subsumes all *certain* constraints, whereas all *uncertain* constraints explicitly depending on u are handled by the inequality $g(x, u) \in -K$. As already mentioned, even if problem (P_u) is well-posed, stable solutions might still become infeasible if the parameter u is only slightly changed: usually the constraint $g(x^*(u), u)$ is active in the optimal solution, i.e. $g(x^*(u), u) \in -K$ but $g(x^*(u), u) \notin \text{int}(-K)$ and hence $g(x^*(u), v) \notin -K$ for some $v \neq u$.

2.1 Local robust counterpart. The *robust counterpart* of the family of uncertain optimization problems (P_u) was introduced by Ben-Tal and Nemirovski in [3] to address this problem of infeasibility under perturbations. For this purpose, the robust counterpart robustifies against perturbations inside some (non-empty, convex and compact) *local uncertainty set* $U \in \mathcal{K}(\bar{U})$ around u :

$$\begin{aligned} \min_{x \in X} \quad & \max_{v \in U} f(x, v) \\ \text{s.t.} \quad & g(x, v) \in -K \quad \forall v \in U. \end{aligned} \quad (RC_U)$$

The robust counterpart approach as such represents a worst-case approach by optimizing the objective over the whole set U of possible parameter realizations. Further, the solution has to be feasible for all realizations of the uncertain parameter within the uncertainty set U . Thus, the feasibility of the solution is immunized against perturbations of the parameter. For more details on robust optimization, see the extensive book by Ben-Tal, El Ghaoui and Nemirovski [2] and the references therein.

To avoid pathological situations, let us assume that there exists at least one feasible point for all uncertain parameters, i.e. there exists an $x_r \in X$ with $g(x_r, u) \in -K$ for all $u \in \bar{U}$.

2.2 Problem reformulation. In the above paragraph, the robust counterpart (RC_U) has been introduced as a convex semi-infinite conic problem. For convenience of the proofs later on let us derive an equivalent convex reformulation of both (P_u) and (RC_U) with only one real-valued inequality constraint. Although this comes at the price of non-smoothness of the constraint, this allows an easy analysis of the consistency of both the original conic problem as well as the semi-infinite conic robust counterpart. As we did not assume smoothness in (P_u), loss of differentiability of the constraint g is thus not of great importance in the following.

For the purpose of reformulation, let us introduce the function $\tilde{g}_z : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}$ for a given constraint g . The structure of \tilde{g}_z is based on ideas in Gomez and Gomez [14] and Werner [37]:

$$\tilde{g}_z(x, u) := \max_{\substack{\lambda \in K^* \\ z^\top \lambda = 1}} \lambda^\top g(x, u), \quad (2.1)$$

where the *positive anchor* $z \in \text{int}(K)$ is chosen arbitrarily, but fixed and K^* denotes the dual cone. In Gomez and Gomez [14] it was shown that the convex feasibility set $\Lambda := \{\lambda \in K^* \mid z^\top \lambda = 1\}$ is indeed compact and thus \tilde{g}_z is well-defined (i.e. finite everywhere). It can easily be shown that \tilde{g}_z inherits all desired properties from the original function g .

PROPOSITION 2.1 *Under the assumptions of Section 1.2 it holds that*

- (i) $\tilde{g}_z(x, u)$ is continuous in (x, u) and convex in x for all $u \in \bar{U}$.
- (ii) $\{x \in X \mid g(x, u) \in -K\} = \{x \in X \mid \tilde{g}_z(x, u) \leq 0\}$ for all $u \in \bar{U}$.
- (iii) $\{x \in X \mid g(x, u) \in \text{int}(-K)\} = \{x \in X \mid \tilde{g}_z(x, u) < 0\}$ for all $u \in \bar{U}$.
- (iv) $\{x \in X \mid g(x, u) \in -K, \forall u \in U\} = \{x \in X \mid \tilde{g}_z(x, u) \leq 0, \forall u \in U\}$ for all $U \in \mathcal{K}(\bar{U})$.
- (v) $\{x \in X \mid g(x, u) \in \text{int}(-K), \forall u \in U\} = \{x \in X \mid \tilde{g}_z(x, u) < 0, \forall u \in U\}$ for all $U \in \mathcal{K}(\bar{U})$.

From the above proposition, it can be deduced that the constraints g and \tilde{g}_z describe the same feasibility set and that also the set of Slater points remains the same. This means that problem (P_u) can be equivalently reformulated to the convex problem

$$\begin{aligned} \min_{x \in X} \quad & f(x, u) \\ \text{s.t.} \quad & \tilde{g}_z(x, u) \leq 0 \end{aligned} \quad (P'_u)$$

with only one real-valued continuous convex constraint. Further, due to Proposition 2.1(iv) and (v) the robust counterpart (RC_U) can be equivalently rewritten as

$$\begin{aligned} \min_{x \in X} \quad & \max_{v \in U} f(x, v) \\ \text{s.t.} \quad & \tilde{g}_z(x, v) \leq 0 \quad \forall v \in U. \end{aligned} \quad (RC'_U)$$

which coincides with the robust counterpart of (P'_u). A better understanding of the reformulation may be gained from the following equality (see Werner [37] for a detailed proof of the equality and for more details on the interpretation),

$$\tilde{g}_z(x, u) = \max_{\substack{\lambda \in K^* \\ z^\top \lambda = 1}} \lambda^\top g(x, u) = \min_{\substack{\nu \in \mathbb{R} \\ g(x, u) - \nu z \in -K}} \nu,$$

which leads to the interpretation of $\tilde{g}_z(x, u)$ as the distance of $g(x, u)$ to the boundary of (in-)feasibility. As shown in Werner [37], these distances lead to tractable formulations for LPs, SOCPs and SDPs if the anchor is properly chosen.

PROOF OF PROPOSITION 2.1. Continuity of \tilde{g}_z in (x, u) follows directly from Theorem 4.3.3 in Bank et al [1]. Further, due to K -convexity of $g(x, u)$ in x for all $u \in \bar{U}$, the mapping $x \mapsto \lambda^\top g(x, u)$ is convex in x in the usual sense for all $u \in \bar{U}$. As Λ is compact, \tilde{g}_z is finite and convex as the pointwise maximum of convex functions, see e.g. Rockafellar [32], Theorem 5.5, thus statement (i) holds. Statements (ii) and (iii) follow directly from the equivalences

$$a \in -K \iff b^\top a \leq 0 \quad \forall b \in K^*$$

and

$$a \in \text{int}(-K) \iff b^\top a < 0 \quad \forall b \in K^*, b \neq 0 \iff b^\top a < 0 \quad \forall b \in K^*, b^\top z = 1.$$

see e.g. Luenberger [22], Proposition 1, p. 215. Finally, statements (iv) and (v) follow from (ii) and (iii) and the compactness of U . \square

2.3 The robust counterpart as a parametric problem. Close inspection of the robust formulation (RC'_U) reveals that (RC'_U) can also be interpreted as a parametric problem, where the role of the parameter $u \in \bar{U}$ is now taken by $U \in \mathcal{K}(\bar{U})$. For this purpose let

$$F(x, U) := \max_{v \in U} f(x, v), \quad \text{and} \quad G(x, U) := \max_{v \in U} \tilde{g}_z(x, v), \quad (2.2)$$

which are well-defined (i.e. finite everywhere). With these definitions, (RC'_U) becomes

$$\begin{aligned} \min_{x \in X} \quad & F(x, U) \\ \text{s.t.} \quad & G(x, U) \leq 0. \end{aligned} \quad (RC'_U)$$

For brevity of notation we will subsequently use Φ and \mathcal{S} (together with F^*) not only for mappings in u , but also for mappings in U as well, if the meaning is clear from the context. However, before the feasible and optimal set mapping for (RC'_U) are investigated in more detail, some additional properties of the objective F and the constraint G of (RC'_U) can be derived. The following proposition shows that the robustified objective function F retains convexity (in x) of the original objective function f . Of course, the same holds true for the constraint G , which inherits convexity of the modified constraint \tilde{g}_z .

Unfortunately, as will be shown in the subsequent Example 2.1, the uniqueness assumption (SA) for the optimal solution of (P_u) does not carry over to (RC_U) , although uniqueness holds for all $u \in \bar{U}$. Nonetheless, the stronger assumption (CA) of strict (quasi-)convexity of the original objective passes on to the robustified version.

THEOREM 2.1 *Let the assumptions of Section 1.2 be satisfied. Then the following statements hold:*

- (i) *The robustified objective and constraint functions F and G are continuous in (x, U) .*
- (ii) *F and G are convex in x for all $U \in \mathcal{K}(\bar{U})$.*
- (iii) *F is strictly (quasi-)convex in x for all $U \in \mathcal{K}(\bar{U})$, if f is strictly (quasi-)convex in x for all $u \in \bar{U}$.*

PROOF. To see statement (i), let us rewrite the objective function F as

$$F(x, U) = \max_{v \in \mathcal{F}(x, U)} h(v, (x, U)) \quad \text{with } \mathcal{F}(x, U) = U \quad \text{and} \quad h(v, (x, U)) = f(x, v).$$

Interpreting (x, U) as a parameter, we immediately see that \mathcal{F} is Hausdorff continuous in this parameter and that further the objective function is continuous in this parameter. In addition, the objective function h is continuous in the optimization variable v and the feasible set $\mathcal{F}(x, U) = U$ is compact. Hence we can apply Theorem 4.2.2(1) and (2) of Bank et al [1], from which the statement follows. For the constraint G , the same reasoning holds. Statement (ii) can be shown in the same manner as in the proof of Proposition 2.1. Finally, statement (iii) follows from the following considerations: Let $0 < \lambda < 1$, $x \neq y$ and $v \in \bar{U}$ arbitrary. Then

$$f(\lambda x + (1 - \lambda)y, v) < \lambda f(x, v) + (1 - \lambda)f(y, v)$$

due to strict convexity of f . Due to continuity of f and compactness of U , taking the maximum over all $v \in U$ yields

$$\begin{aligned} \max_{v \in U} f(\lambda x + (1 - \lambda)y, v) &< \max_{v \in U} \lambda f(x, v) + (1 - \lambda)f(y, v) \\ &\leq \max_{v \in U} \lambda f(x, v) + \max_{v \in U} (1 - \lambda)f(y, v) \end{aligned}$$

and thus

$$F(\lambda x + (1 - \lambda)y, U) < \lambda F(x, U) + (1 - \lambda)F(y, U)$$

which is strict convexity of the robustified objective F . Similar arguments hold for the quasiconvex case.

□

2.4 Well-posedness of (RC_U) . We note that the robust counterpart (RC_U) generalizes (P_u) in the sense that it coincides with the original problem if $U = \{u\}$. From the structure of F and G , we further expect that the robust counterpart inherits most properties of the original problem, at least for sufficiently small local uncertainty sets U . As shown in the next two theorems (Theorem 2.2 and Theorem 2.3), this is indeed the case, i.e. existence and stability of solutions carry over from (P_u) to (RC_U) under reasonably mild assumptions on the local uncertainty set. We will further see that this is the key result to establish an expected connection between the original and the robust formulation – that the solution of the robust counterpart converges to the optimal solution of the original problem for shrinking local uncertainty sets.

Unfortunately, as already mentioned, the uniqueness assumption of the original problem (P_u) does not necessarily hold for (RC_U) as well. The following example has been constructed in such a way that f is strictly convex for almost all $u \in \bar{U} = [0; 1]$ besides the corner point $u = 1$. Still, $\mathcal{S}(1)$ is a singleton due to construction, i.e. the uniqueness assumption still holds. However, as we will see, the robust counterpart does not possess a unique solution for any $U = [t; 1]$ with $0 \leq t < 1$ arbitrary.

EXAMPLE 2.1 *Let us consider the following two-dimensional problem, which fulfills all requirements from Section 1.2, the uniqueness assumption (SA), the Slater condition (SC_u) for all $u \in \bar{U}$ and all assumptions from Section 2.1:*

$$\begin{aligned} \min_{x \in [0; 1]^2} \quad & f(x, u) := u \cdot \|x\|_1 + (1 - u) \cdot \|x\|_2^2 \\ \text{s.t.} \quad & g(x, u) := (1 - u) - (x_1 + x_2) \leq 0. \end{aligned}$$

Now, let $\bar{U} = [0; 1]$ and $U_t = [t; 1]$ with $0 \leq t < 1$ arbitrary. Obviously, for all $u \in [0; 1[$ the objective is strictly convex and hence the optimal solution is unique. For $u = 1$, the objective becomes $f(x, 1) = \|x\|_1$ which is not strictly quasiconvex. Nevertheless, in this case, the optimal solution $x^*(1) = (0, 0)^\top$ is unique, as $(0, 0)^\top$ satisfies the constraint $g(x, 1) = -(x_1 + x_2) \leq 0$, hence assumption (SA) is fulfilled.

Using $U_t = [t; 1]$, the robust objective function F becomes $F(x, U_t) = f(x, 1) = \|x\|_1$, as $\|x\|_1 \geq \|x\|_2^2$ on $[0; 1]^2$. Additionally, the robust constraint G becomes $G(x, U_t) = g(x, t) = (1 - t) - (x_1 + x_2) \leq 0$. Thus, for fixed t all points $x = (z, (1 - t) - z)^\top$ are optimal for $0 \leq z \leq 1 - t$, and the solution of the robust counterpart is not unique.

From the above example it can be deduced that either some stronger (strict quasiconvexity) assumption than uniqueness has to be imposed on (P_u) , or the uniqueness assumption has to be imposed for (RC_U) as well.

THEOREM 2.2 *Under the assumptions from Section 1.2 and Section 2.1, problem (RC_U) possesses at least one optimal solution for all local uncertainty sets $U \in \mathcal{K}(\bar{U})$. If further (CA) holds, then $\mathcal{S}(U)$ is a singleton.*

PROOF. First of all, we note that each problem possesses at least one feasible point $x_r \in \Phi(U)$ due to the assumption from Section 2.1. Further, the feasible sets $\Phi(U)$ are convex and compact due to continuity and convexity of G (see Proposition 2.1(i) and (ii)) and convexity and compactness of X . This already guarantees existence of at least one optimal solution. Finally, due to Proposition 2.1(iii), F is strictly quasiconvex in x if f is strictly quasiconvex, hence the optimal solution is unique. \square

So far we have not defined what we understand as a sufficiently small local uncertainty set. For the following, we therefore introduce the notion of a δ -small local uncertainty set.

DEFINITION 2.1 *A local uncertainty set $U \in \mathcal{K}(\bar{U})$ is called δ -small around $\bar{u} \in \bar{U}$ if $U \subset B_\delta(\bar{u}) \cap \bar{U}$ where $B_\delta(\bar{u})$ denotes the closed δ -ball around \bar{u} . A local uncertainty set U is called δ -small if there exists an $\bar{u} \in \bar{U}$ such that U is δ -small around \bar{u} .*

LEMMA 2.1 *Additional to the assumptions in Sections 1.2 and 2.1, we further assume that the Slater condition (SC_u) holds for (P_u) for some $u \in \bar{U}$. Then there exists a small enough $\delta = \delta(u) > 0$ such that the Slater condition (SC_U) holds for (RC_U) for all δ -small local uncertainty sets U around u , i.e. it holds*

$$\exists x_{Sl}(U) \in X \text{ such that } G(x_{Sl}(U), U) < 0. \quad (SC_U)$$

PROOF. As $G(x, U_1) \leq G(x, U_2)$ for all $x \in X$ and all $U_1 \subset U_2$, it is sufficient to prove the statement only for local uncertainty sets $U = B_\delta(u)$. Let us choose a sequence $\delta_n \rightarrow 0$, then it obviously holds for $U_n := B_{\delta_n}(u)$ that $U_n \xrightarrow{H} B_0(u) = \{u\}$. Due to Proposition 2.1(i), G is continuous in U , hence it follows that $G(x_{Sl}(u), U_n) \rightarrow G(x_{Sl}(u), \{u\}) = \tilde{g}_z(x_{Sl}(u), u) < 0$ as assumed. Therefore, $x_{Sl}(u)$ remains a Slater point for the robustified problem, if the uncertainty set is sufficiently small. \square

In the above lemma, the smallness parameter δ is still depending on the uncertain data u . The following result, which in idea is similar to Bliman and Prieur [5], shows that in fact δ can be chosen uniformly, i.e. independent of u .

PROPOSITION 2.2 *Additional to the assumptions in Sections 1.2 and 2.1, we further assume that (SC_u) holds for (P_u) for all $u \in \bar{U}$. Then there exists a small enough $\delta > 0$ such that (SC_U) holds for all δ -small local uncertainty sets U .*

PROOF. For $k \in \mathbb{N}$ let us introduce the set

$$Z_k := \{u \in \bar{U} \mid \exists x \in X : G(x, B_{\frac{1}{k}}(u)) < 0\}$$

of all parameters u where a Slater point can be found for a smallness level of $\delta_k = \frac{1}{k}$. Obviously $Z_k \subset Z_{k+1} \subset \bar{U}$ for all k . The statement of the proposition is thus equivalent to the existence of a $k_0 \in \mathbb{N}$ with $Z_{k_0} = U$. Then, δ could be chosen as $\delta_{k_0} = \frac{1}{k_0}$.

Assume that for all k , the set Z_k is not equal to \bar{U} . Then there exists a sequence $(u_k)_k$ with $u_k \in \bar{U} \setminus Z_k$. Since \bar{U} is compact, this sequence has an accumulation point in \bar{U} , say \bar{u} , and without loss of generality $\bar{u} = \lim_{k \rightarrow \infty} u_k$. By assumption, problem $(P_{\bar{u}})$ has a Slater point, i.e. there exists $x_{Sl}(\bar{u})$ such that

$$G(x_{Sl}(\bar{u}), \{\bar{u}\}) = \tilde{g}_z(x_{Sl}(\bar{u}), \bar{u}) < 0$$

As G is continuous in the second argument and $B_{\frac{1}{k}}(u_k) \xrightarrow{H} \{\bar{u}\}$, we have

$$\lim_{k \rightarrow \infty} G(x_{Sl}(\bar{u}), B_{\frac{1}{k}}(u_k)) = G(x_{Sl}(\bar{u}), \{\bar{u}\}) < 0$$

which shows that for sufficiently large k , $x_{Sl}(\bar{u})$ is a Slater point for u_k at smallness level $\frac{1}{k}$, i.e. $u_k \in Z_k$, which is a contradiction. \square

Now we are ready to prove the main result on the well-posedness of the robust counterpart, which is the analogue to Theorem 1.2 for the original problem.

THEOREM 2.3 *Let the assumptions of Sections 1.2 and 2.1 be satisfied. If the Slater condition (SC_u) holds for $u \in \bar{U}$ then there exists a sufficiently small $\delta = \delta(u) > 0$ such that*

- (i) *the feasible set mapping Φ is Hausdorff continuous in U ,*
- (ii) *the optimal value function f^* is continuous in U , and*
- (iii) *the optimal set mapping \mathcal{S} is Hausdorff upper semicontinuous in U*

for all δ -small local uncertainty sets $U \in \mathcal{K}(\bar{U})$ around u .

PROOF. Based on the reformulation (RC'_U) of (RC_U) , the proof of (i) and (ii) remains the same as for Theorem 1.2, with Hausdorff continuity replacing ordinary continuity where necessary. Statement (iii) follows from Proposition 1.1. \square

COROLLARY 2.1 *Let the assumptions from Theorem 2.3 be satisfied for some $u \in \bar{U}$. Then for each sequence of local uncertainty sets U_n with $U_n \xrightarrow{H} \{u\}$ it holds that each cluster point of the sequence of corresponding solutions $x^*(U_n)$ is optimal in (P_u) .*

Based on the previous theorem and the above corollary, the results by Ben-Tal and Nemirovski (see [3], Theorem 2.1 and Section 2.3) follow as an easy consequence: the optimal value of the robust counterpart converges to the optimal value of the original problem, if the uncertainty is shrunk to a singleton. It has to be noted that this result holds in a broader context than the original results by Ben-Tal and Nemirovski, however, even for LPs, a Slater condition needs to be satisfied.

COROLLARY 2.2 *Let the assumptions from Theorem 2.3 be satisfied for all $u \in \bar{U}$ and let $\mathcal{S}(U)$ be a singleton for all $U \in \mathcal{K}(\bar{U})$. Then there exists a sufficiently small $\delta > 0$ such that problem (RC_U) is well-posed for all δ -small local uncertainty sets U , i.e. the mapping $U \mapsto x^*(U)$ is continuous in U .*

From the preceding corollary, the already stated connection between the original problem and the robust counterpart can be established:

COROLLARY 2.3 *Under the assumptions of Corollary 2.2, it holds that $x^*(U_k) \rightarrow x^*(u)$ for all sequences $(U_k)_k$ of local uncertainty sets with $U_k \xrightarrow{H} \{u\}$.*

PROOF. As $U_k \xrightarrow{H} \{u\}$, U_k becomes δ -small for sufficiently large k , the statement follows from the continuity of the optimal set mapping, see Corollary 2.2. \square

2.5 Example. Although Corollary 2.1 might seem to be a straightforward result and one would expect this connection to hold in a very general fashion, the next example shows that this is unfortunately not the case. This example is constructed in such a way that the optimal set mapping in the original problem is not Hausdorff upper semicontinuous in $u = 0$ and therefore, the robust solution fails to converge to the true solution. Please note that besides the missing Slater condition, all other assumptions are satisfied.

EXAMPLE 2.2 *Consider the following (strictly convex) problem*

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

with $u \in \bar{U} = [-1, 1]$. It can be noted immediately that the Slater condition is fulfilled for all $u \in \bar{U}$ besides $u = 0$. The corresponding feasible set mapping, optimal value function and optimal set mapping are given by

$$\Phi(u) = \begin{cases} [0; 1] & \text{if } u < 0 \\ [-1; 1] & \text{if } u = 0 \\ [-1; 0] & \text{if } u > 0 \end{cases} \quad f^*(u) = \begin{cases} 0 & \text{if } u < 0 \\ 0 & \text{if } u = 0 \\ 1 & \text{if } u > 0 \end{cases} \quad \mathcal{S}(u) = x^*(u) = \begin{cases} 1 & \text{if } u < 0 \\ 1 & \text{if } u = 0 \\ 0 & \text{if } u > 0. \end{cases}$$

It can be observed that, as expected, Φ is Hausdorff upper semicontinuous at $u = 0$, but not Hausdorff lower semicontinuous. Further, neither f^* nor x^* are continuous in $u = 0$.

Now let us consider the corresponding robust counterpart with local uncertainty set $U = B_\delta(u) = [-\delta + u, \delta + u]$. Then the following reformulation of the robust counterpart can be easily obtained:

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

The optimal solution thereof depends on the center u and the size δ of the uncertainty set and it is given by

$$\mathcal{S}(U) = x^*(U) = x^*(B_\delta(u)) = \begin{cases} 1 & \text{if } u < -\delta \\ 0 & \text{else.} \end{cases}$$

Interestingly, the robust solution x^* is locally Hausdorff continuous for all local uncertainty sets $U = B_\delta(u)$ with $u \neq -\delta$. This means especially that for all $u \neq 0$, the robust solution converges to the original solution for $\delta \rightarrow 0$ as expected by Corollary 2.3. However, for $u = 0$, the robust solution $x^*(B_\delta(0)) = 0$ does not converge to the original optimal solution $x^*(0) = 1$. In other words, neither the original nor the robust solution is Hausdorff continuous at $u = 0$.

3. Consistency of robust optimization. In the case that the uncertain data is obtained by a statistical estimation procedure, the whole framework of statistical theory can be applied. In the following, it will be recalled that the *plug-in estimator*

$$\hat{x} := x^*(\hat{u}) \in \mathcal{S}(\hat{u})$$

obtained as solution of $(P_{\hat{u}})$ is consistent if the estimator \hat{u} is already consistent. Furthermore, it will be demonstrated that the plug-in estimator can easily be supplemented by another consistent estimator, namely the *robustified plug-in estimator*

$$\hat{x}_{\text{robust}} := x^*(\hat{U}) \in \mathcal{S}(\hat{U}),$$

determined as solution of the corresponding robust counterpart problem with uncertainty set \hat{U} satisfying some additional properties.

3.1 Consistent point estimators. In the next paragraphs we will briefly recall the main definitions and properties of *point estimators* for some unknown parameter. For this purpose, let a suitable probability space $(\Omega, \mathcal{A}, \mathbf{P})$ be given, together with a collection of \mathbb{R}^{d_1} -valued random variables Z_1, \dots, Z_N , where Z_i , $i = 1, \dots, N$, are independent and identically distributed. Assume further that the distribution $\mathbf{P}_{Z_i} = \mathbf{P}_\theta$ is depending on some unknown parameter θ which needs to be estimated. In this case any *statistic* $t : (\mathbb{R}^{d_1})^N \rightarrow \mathbb{R}^{d_2}$, i.e. any measurable function t defined on samples (z_1, \dots, z_N) , yields a corresponding *point estimator* $\hat{\theta}$ for the unknown parameter θ by $\hat{\theta} = t(Z_1, \dots, Z_N)$. For more details on statistics and point estimators, we refer to Müller [27] or Rinne [31].

In general, although it is not possible to characterize the distribution of an arbitrary point estimator $\hat{\theta}$ in analytical terms, the most relevant properties of point estimators can still be characterized:

- A point estimator $\hat{\theta}$ is called *unbiased* if $\mathbf{E}_\theta[\hat{\theta}] := \mathbf{E}_{\mathbf{P}_\theta}[\hat{\theta}] = \theta$.
- A sequence of point estimators $(\hat{\theta}_N)_{N \in \mathbb{N}}$ is called *asymptotically unbiased* if $\mathbf{E}_\theta[\hat{\theta}_N] \rightarrow \theta$.
- A sequence of asymptotically unbiased estimators is called
 - *consistent* if $\hat{\theta}_N \xrightarrow{P} \theta$, i.e. $\mathbf{P}[\{\omega \in \Omega \mid \|\hat{\theta}_N - \theta\| > \varepsilon\}] \rightarrow 0$ for all $\varepsilon > 0$, and
 - *strongly consistent* if $\hat{\theta}_N \xrightarrow{a.s.} \theta$, i.e. $\mathbf{P}[\{\omega \in \Omega \mid \hat{\theta}_N \rightarrow \theta\}] = 1$,

where the convergence of the estimators $\hat{\theta}_N$ is known as convergence in probability and almost sure convergence, respectively.

- A sequence of point estimators $\hat{\theta}_N$ is called *asymptotically normally distributed* with asymptotic covariance matrix K , if there exists a random variable $\epsilon \sim \mathcal{N}(0, K)$ such that

$$\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} \epsilon.$$

where the convergence is understood in distribution.

For more details on the different types of convergence and their relationship, we refer to Breiman [8].

Unfortunately, an unbiased point estimator does not remain unbiased under continuous nonlinear transformations. In terms of our previous considerations this means that if we are given an unbiased point estimator \hat{u} for some parameter $u \in \bar{U}$ (i.e. $\mathbf{E}_u[\hat{u}] = u$), then the plug-in estimator \hat{x} is usually not unbiased, i.e. it usually holds that $\mathbf{E}_u[\hat{x}] = \mathbf{E}_u[x^*(\hat{u})] \neq x^*(u)$. From this point of view, it does not seem advisable to calculate an estimator for the true optimal solution $x^*(u)$ by such a simple plug-in rule. However, this bias vanishes asymptotically, as can be seen from the next two theorems, which are a direct consequence of the Continuous Mapping Theorem (also known as Mann-Wald-Theorem) in probability theory.

THEOREM 3.1 *If the sequence of estimators $\hat{\theta}_N$ is (strongly) consistent for θ and the function h is continuous in θ , then the sequence of estimators $h(\hat{\theta}_N)$ is (strongly) consistent for $h(\theta)$.*

PROOF. See Billingsley [4]. □

From the preceding Theorem 3.1, we can immediately deduce the following theorem, which is especially tailored to the particular situation of this exposition.

THEOREM 3.2 *Let the sequence of estimators \hat{u}_N be (strongly) consistent for a parameter $u \in \bar{U}$ and let the assumptions of Theorem 1.2 be satisfied, i.e. let (P_u) be well-posed. Then the sequence of plug-in estimators $\hat{x}_N = x^*(\hat{u}_N)$ is (strongly) consistent for $x^*(u)$.*

PROOF. This follows directly from Theorems 1.2 and 3.1. □

3.2 Examples and remarks on consistent estimators. Let us give an important example for consistent estimators, which play a prominent role in portfolio optimization (see Section 4) before we state some remarks on alternative, equally popular robust estimators.

EXAMPLE 3.1 Let $R \in \mathbb{R}^n$, $R \sim \mathcal{N}(\mu, \Sigma)$ and let N independent realizations r_1, \dots, r_N be given. Then the maximum likelihood estimators for μ and Σ are given by

$$\hat{\mu}_N := \frac{1}{N} \sum_{k=1}^N r_k, \quad \text{and} \quad \hat{\Sigma}_N := \frac{1}{N} \sum_{k=1}^N (r_k - \hat{\mu}_N)(r_k - \hat{\mu}_N)^\top. \quad (3.1)$$

In this special case the joint distribution of the maximum likelihood estimators $\hat{\mu}$ and $\hat{\Sigma}$ is known to be the product of a normal and a Wishart distribution, see for example Press [30], Section 7.1:

$$(\hat{\mu}_N, \hat{\Sigma}_N) \sim \mathcal{N}\left(\mu, \frac{1}{N} \Sigma\right) \otimes \mathcal{W}\left(\frac{1}{N} \Sigma, N-1\right), \quad (3.2)$$

where $\mathcal{W}(Z, q)$ denotes the Wishart distribution with scale matrix $Z \in \mathbb{R}^{n \times n}$ and q degrees of freedom. From Equation (3.2) it can be easily concluded that both $\hat{\mu}$ and $\hat{\Sigma}$ are strongly consistent and asymptotically normal estimators. Note that although $\hat{\mu}$ is already unbiased, $\hat{\Sigma}$ needs to be scaled by $\frac{N}{N-1}$ to avoid bias.

The above observation concerning consistency and asymptotic normality remains almost always true, i.e. maximum likelihood estimators are known to be consistent under very weak assumptions. Under slightly stronger conditions, maximum likelihood estimators are also strongly consistent and asymptotically normal. These properties presumably put maximum likelihood estimators amongst the most convenient and most favoured estimators, cf. Rinne [31].

Nevertheless, although maximum likelihood estimators probably represent one of the most popular point estimators for the mean μ , other estimators, for example *L-estimators* like the trimmed mean, the median or some percentile estimator, or alternatively *M-estimators* as the Huber estimator are sometimes preferred in practical applications due to their improved robustness properties, see e.g. Huber [16]. Again, it is known that under rather weak conditions, both L-estimators and M-estimators are unbiased and asymptotically normal (see e.g. Rinne [31] as well as Huber [16]), and thus consistent.

3.3 Consistency of robust optimization. As seen in Theorem 3.2, consistent estimators \hat{x} for some optimal solution $x^*(u)$ can be obtained by plugging in any consistent estimator \hat{u} for the data u . Moving from the original problem to the robust counterpart, the natural question arises if the robust solution remains consistent as well, or if this property is lost by robustification. To be able to provide a confirmative answer to this question, we need to consider a stochastic analogy of the Hausdorff convergence of a sequence of sets U_N to a point u where the sets U_N now become random variables. For this purpose we note that the metric space $(\mathcal{K}(\bar{U}), d_H)$ becomes a measure space if it is equipped with the Borel- σ -algebra \mathcal{B}_{d_H} which corresponds to the metric d_H , see for example Doob [10], II.4. Therefore, $\mathcal{K}(\bar{U})$ -valued (i.e. set-valued) random variables are represented by \mathcal{A} - \mathcal{B}_{d_H} -measurable functions. For a random set U the random variable

$$d_H(U, \{u\}) = \max_{v \in U} \|u - v\|$$

which represents the Hausdorff distance of the random set U to the singleton $\{u\}$ is a measurable random variable on the given probability space $(\Omega, \mathcal{A}, \mathbf{P})$, as

$$U \mapsto \max_{v \in U} \|u - v\|$$

is a continuous and thus measurable mapping. With these preliminaries, we are now ready to introduce the following notions of consistency of random uncertainty sets:

DEFINITION 3.1 A sequence of random sets $U_N \in \mathcal{K}(\bar{U})$ is called consistent for $u \in \bar{U}$ if $d_H(U_N, \{u\}) \xrightarrow{P} 0$ for $N \rightarrow \infty$, i.e. if $\mathbf{P}[\{\omega \in \Omega \mid d_H(U_N(\omega), \{u\}) > \varepsilon\}] \rightarrow 0$ for all $\varepsilon > 0$. It is called strongly consistent if $d_H(U_N, \{u\}) \xrightarrow{a.s.} 0$ for $N \rightarrow \infty$, i.e. if $\mathbf{P}[\{\omega \in \Omega \mid U_N(\omega) \xrightarrow{H} \{u\}\}] = 1$.

Based on the notion of consistent uncertainty sets it is now easy to show that the robust solutions provide a consistent estimator in the same way as the solution of the original problem.

THEOREM 3.3 *Let the sequence of uncertainty sets U_N be (strongly) consistent for a parameter $u \in \bar{U}$ and let the assumptions of Corollary 2.2 be satisfied, i.e. let (RC_{U_N}) be well-posed for sufficiently large N . Then the sequence of plug-in estimators $x^*(U_N)$ is (strongly) consistent for $x^*(u)$.*

Note that the proof of this theorem follows directly from the well-posedness of the robust counterpart and the Continuous Mapping Theorem in the same way as for the original problem. However, to avoid confusion with the different kind of convergences, we present a more detailed proof for the case of strong consistency.

PROOF OF THEOREM 3.3. Let $\mathcal{B} := \{\omega \in \Omega \mid U_N(\omega) \xrightarrow{H} \{u\}\}$. Due to strong consistency we have $\mathbf{P}[\mathcal{B}] = 1$. Applying Corollary 2.3 we obtain

$$\mathcal{B} \subset \{\omega \in \Omega \mid x^*(U_N(\omega)) \rightarrow x^*(u)\}$$

and therefore $\mathbf{P}[\{\omega \in \Omega \mid x^*(U_N(\omega)) \rightarrow x^*(u)\}] = 1$, which is exactly almost sure convergence, i.e. strong consistency of $x^*(U_N)$. \square

3.4 Examples of consistent uncertainty sets. Based on the above Example 3.1, we will show in the subsequent example that confidence regions which correspond to consistent point estimators are usually instances of consistent uncertainty sets.

EXAMPLE 3.2 *Let the setup of Example 3.1 be given and define an associated confidence ellipsoid around the point estimator $\hat{\mu}$ by*

$$\begin{aligned} U_N^1 &:= U_{N,\delta}^1(\hat{\mu}_N) := \{r \in \mathbb{R}^n \mid (r - \hat{\mu}_N)^T \left(\frac{1}{N} \hat{\Sigma}_N\right)^{-1} (r - \hat{\mu}_N) - \delta^2 \leq 0\} \\ &= \left\{r \in \mathbb{R}^n \mid N(r - \hat{\mu}_N)^T \hat{\Sigma}_N^{-1} (r - \hat{\mu}_N) - \delta^2 \leq 0\right\} \\ &= \left\{r \in \mathbb{R}^n \mid r = \hat{\mu}_N + \delta \frac{1}{\sqrt{N}} \hat{\Sigma}_N^{\frac{1}{2}} y, \|y\|_2 \leq 1\right\}. \end{aligned} \quad (3.3)$$

The fixed size δ^2 is usually obtained by an appropriate α -quantile, i.e. δ^2 is chosen such that $\alpha = \chi_n^2(\delta^2)$, with $\alpha \in (0, 1)$ representing the desired confidence, because then $\mathbf{P}[\mu \in \mathcal{U}_{N,\delta}^1(\hat{\mu}_N)] = \alpha$, independent of N . From its definition it can easily be observed that U_N^1 is (strongly) consistent for μ , if $\hat{\mu}$ and $\hat{\Sigma}$ are (strongly) consistent point estimators for μ and Σ .

The above uncertainty set U_N^1 is designed in such a way that it covers uncertainty of μ , however, no ambiguity of Σ is considered in its construction. In a more general fashion, a joint uncertainty for μ and Σ will be introduced in the subsequent example.

EXAMPLE 3.3 *Let the setup of Example 3.1 be given and let a joint uncertainty set for (μ, Σ) be defined by*

$$U_N^2 := U_{N,\delta}^2(\hat{\mu}_N, \hat{\Sigma}_N) := \{(r, C) \in \mathbb{R}^n \times \mathbb{S}^n \mid N(r - \hat{\mu}_N)^T \hat{\Sigma}_N^{-1} (r - \hat{\mu}_N) + \frac{N-1}{2} \|\hat{\Sigma}_N^{-\frac{1}{2}} (C - \hat{\Sigma}_N) \hat{\Sigma}_N^{-\frac{1}{2}}\|_{\text{tr}}^2 \leq \delta^2\},$$

where $\|A\|_{\text{tr}}^2 = \text{tr}(A^T A)$. It has been argued in Schöttle and Werner [35], Section 3.2, that U_N^2 is indeed a canonical choice for a joint uncertainty set of $\hat{\mu}$ and $\hat{\Sigma}$. As before, it can easily be shown that U_N^2 is (strongly) consistent for (μ, Σ) if $\hat{\mu}$ and $\hat{\Sigma}$ are (strongly) consistent point estimators for μ and Σ .

It is worth mentioning that most consistent uncertainty sets in practical applications are directly based upon some confidence region, i.e. are based on statistical reasoning. For instance, a variety of uncertainty sets has been investigated in portfolio optimization related literature so far, see for example Ceria and Stubbs [9], Lutgens [23] or Meucci [25] for specific choices of uncertainty sets. In particular, let us mention two approaches which are slightly different from the above: first, Goldfarb and Iyengar [13], Section 5, consider an uncertainty set motivated by a linear factor model and second, Schöttle and Werner [34], Section 2.3, investigate uncertainty sets derived from several different point estimators. In both instances it can be shown that the uncertainty sets are consistent, too.

4. Application to mean-variance portfolio optimization. In this section, we will apply the previous results on well-posedness to the specific problem of mean-variance portfolio optimization. Concerning literature, there is only little work on consistency of portfolio estimators to directly relate to. Instead, most publications focus on the practically more relevant case of finite samples (see for example Okhrin and Schmid [28], Kan and Smith [19] or Kan and Zhou [20] and the references therein), but do not consider the asymptotical statistical properties of the according portfolio estimator. As main exceptions, there is the seminal paper [18] by Jobson and Korkie which provides characteristics (e.g. moments) and asymptotic properties of portfolio estimates, especially asymptotic normality (and thus consistency) of the optimal portfolio. However, their results are all based on the assumption of normally distributed returns as well as the specific choice of the maximum likelihood estimator and are further strongly dependent on the fact that no constraints on the portfolios are considered and that hence an analytical formula for the optimal portfolio is available. The only other relevant work on asymptotic properties of portfolio estimators is Mori [26] who generalized the results of Jobson and Korkie to the case of linear (parameter-independent) equality constraints on the portfolio weights. Although Mori also considers a Stein-type estimator in addition to the maximum likelihood estimator, his results are also based on the same assumptions (besides slightly more general portfolio constraints) as in Jobson and Korkie [18] and are again derived from an analytical formula for the optimal portfolio.

On the other hand, there is vast literature on the question how to robustify portfolio estimators. For example, in the paper by Perret-Gentil and Victoria-Feser [29] robustification is obtained by using a robust plug-in estimator. As main result it is shown that for finite samples the stability of optimal portfolios can be improved by such robust plug-in estimators. Alternative ways of robustification based on the robust counterpart have also been successfully considered by several authors, among those are for example Goldfarb and Iyengar [13], Tütüncü and Koenig [36], Ceria and Stubbs [9], Lutgens [23], as well as Schöttle and Werner [34]. All these approaches rely on applications of the robust counterpart methodology for standard optimization problems under uncertainty with appropriately chosen uncertainty sets.

In the subsequent Theorems 4.1 and 4.2, all the setups mentioned in the above paragraphs will be covered in a unified way and consistency of optimal (robust) portfolio estimators is obtained under very general assumptions:

- asset returns are not restricted to normal distributions but are allowed to follow any multivariate elliptical distribution,
- any consistent estimator can be plugged into the mean-variance portfolio, but especially any consistent robust estimator,
- both the ordinary mean-variance problem as well as its robust counterpart are covered, and,
- as noted in Remark 4.1, very general constraints – even depending on the estimators – can be considered as long as a regularity condition is satisfied.

4.1 Mean-variance portfolio optimization setup. We consider a financial market with n risky assets defined on the probability space $(\Omega, \mathcal{A}, \mathbf{P})$, and we restrict ourselves to the single period setting from some time t_0 to t_1 . It is well-known, see for example Ingersoll [17], that in this setting the classical mean-variance theory developed by Markowitz [24] is consistent with utility theory if the asset returns follow a multivariate elliptical distribution. Therefore, assume that the linear one period return $R \in \mathbb{R}^n$ is multivariate elliptically distributed with parameters μ , Σ and characteristic generator ψ , i.e.

$$R \sim \mathcal{E}(\mu, \Sigma, \psi).$$

Further, assume that R possesses a density and has finite second moments, and that $2\psi'(0) = -1$, which means

$$\mathbf{E}[R] = \mu \quad \text{and} \quad \mathbf{Cov}[R] = -2\psi'(0) \Sigma = \Sigma.$$

For more details on elliptical distributions, we refer to Fang and Zhang [12] or Fang et al. [11]. Assume that the set of feasible portfolios X is non-empty, convex and compact and that portfolio weights add up to one, i.e.

$$X \subset \{x \in \mathbb{R}^n \mid x^\top \mathbf{1} = 1\}.$$

This setting covers especially the case considered by Jobson and Korkie as well as the linearly constrained setup of Mori. Compactness is not a very strong restriction, since it is usually fulfilled in practical settings, for example if some short-selling constraints $x \geq x_0$ are present.

If asset returns R are distributed with mean μ and covariance matrix Σ , the expected return of a portfolio $x \in X$ equals $x^\top \mu$ and the corresponding risk figure in the mean-variance framework is given by $\sqrt{x^\top \Sigma x}$. In the traditional mean-variance setup according to Markowitz, an optimal trade-off – represented by a trade-off factor $0 \leq \lambda \leq 1$ – between risk and expected return is sought for:

$$\min_{x \in X} (1 - \lambda) \sqrt{x^\top \Sigma x} - \lambda (x^\top \mu). \quad (MV_{\mu, \Sigma}^{(\lambda)})$$

For a given risk-return trade-off parameter λ , each optimal solution $x^*(\lambda, \mu, \Sigma)$ of problem $(MV_{\mu, \Sigma}^{(\lambda)})$ is a mean-variance *efficient portfolio* in the sense of Markowitz. Using the parameter λ , the efficient portfolios on the *efficient frontier* can be traced from the minimum variance portfolio ($\lambda = 0$) to the maximum return portfolio ($\lambda = 1$). Note that this formulation is completely equivalent to the classical textbook formulation, see e.g. Schöttle and Werner [35] for a detailed proof.

4.2 Consistency of the plug-in estimator in mean-variance portfolio optimization. For each choice of market parameters (μ, Σ) with positive definite matrix Σ , problem $(MV_{\mu, \Sigma}^{(\lambda)})$ is a convex optimization problem for all $0 \leq \lambda \leq 1$ and it possesses at least one optimal solution due to the compactness of X , which in fact is only crucial for $\lambda = 1$, as otherwise existence is guaranteed by the coerciveness of the objective function. It is further easy to see that for $0 \leq \lambda < 1$ the solution to $(MV_{\mu, \Sigma}^{(\lambda)})$ is unique due to strict convexity. For $\lambda = 1$ there might be more than one optimal solution for very specific choices of μ as $(MV_{\mu, \Sigma}^{(\lambda)})$ might become an LP. In the following, let us therefore assume – in accordance with the uniqueness assumption (SA) – that $(MV_{\mu, \Sigma}^{(1)})$ possesses exactly one optimal solution.

Similar to the general setup, if the data – in this case the parameters μ and Σ – is not known, estimators for the unknown quantities have to be used instead. In accordance with the last section (cf. Example 3.1), we assume that these estimators stem from a historical data sample of N historical (independent and identically distributed) realizations r_1, \dots, r_N of R . It can be shown (see Fang and Zhang [12]) that the estimators defined in Equation (3.1) are indeed strongly consistent for all multivariate elliptical distributions. Let us point out that this assumed setup is not restricted to maximum likelihood estimators, but is much more general, as it also covers robust estimators as mentioned in Section 3.2.

Now, continuity of the optimal solution mapping $(\mu, \Sigma) \mapsto x^*(\lambda, \mu, \Sigma)$ follows immediately by applying Theorem 1.2:

COROLLARY 4.1 *Under the assumptions of Section 4.1 it holds that the optimal solution mapping $(\mu, \Sigma) \mapsto x^*(\lambda, \mu, \Sigma)$ is continuous on $\mathbb{R}^n \times \text{int}(\mathbf{S}_+^n)$ for all $0 \leq \lambda \leq 1$.*

PROOF. We first observe that the feasible set X in $MV_{\mu, \Sigma}^{(\lambda)}$ is compact and independent of (μ, Σ) . Therefore, the Slater condition which is assumed in Theorem 1.2 can be dropped. As the remaining assumptions of Theorem 1.2 are satisfied, problem $(MV_{\mu, \Sigma}^{(\lambda)})$ is well-posed. \square

REMARK 4.1 *According to Theorem 1.2, much more general problem formulations than $(MV_{\mu, \Sigma}^{(\lambda)})$ could be considered, which still have continuous optimal set mappings. Obviously, additional constraints could be added, as long as a Slater condition is satisfied. Further, it is also possible to allow that X depends Hausdorff continuously on (μ, Σ) .*

From Corollary 4.1 it immediately follows that the plug-in portfolio estimator \hat{x} is a strongly consistent estimator for the true optimal portfolio.

THEOREM 4.1 *Let $0 \leq \lambda \leq 1$ and let the assumptions of Section 4.1 be satisfied. Then the plug-in portfolio estimator $\hat{x}_N = x^*(\lambda, \hat{\mu}_N, \hat{\Sigma}_N)$ is a (strongly) consistent estimator for the true optimal portfolio $x^*(\lambda, \mu, \Sigma)$ if the estimators $\hat{\mu}_N$ and $\hat{\Sigma}_N$ are (strongly) consistent estimators for μ and Σ .*

PROOF. Follows immediately from Corollary 4.1 together with Theorem 3.2. \square

Although the consistency result stated in Theorem 4.1 seems to be derived in an easy manner, let us again point out that this result actually generalizes all previous results on consistency by Jobson and by Mori.

4.3 Consistency of robust mean-variance portfolio estimators. In the same way as consistency of the original mean-variance portfolios has been derived, it is possible to extend consistency to robust portfolios, i.e. to solutions $x^*(\lambda, U)$ of the robust portfolio problem

$$\min_{x \in X} \max_{(r, C) \in U} (1 - \lambda) \sqrt{x^\top C x} - \lambda(x^\top r). \quad (RMV_U^{(\lambda)})$$

For this purpose, consistent uncertainty sets replace consistent point estimators, as motivated in the previous section.

THEOREM 4.2 *Let $0 \leq \lambda < 1$ and let the assumptions of Section 4.1 be satisfied. Then the robust portfolio estimator $\hat{x}_N = x^*(\lambda, U_N)$ is a (strongly) consistent estimator for the true optimal portfolio $x^*(\lambda, \mu, \Sigma)$ if the consistency sets U_N are (strongly) consistent for (μ, Σ) .*

PROOF. As U is convex and compact, and as the objective function of the original mean-variance problem is strictly convex for $\lambda < 1$, we can apply Theorem 2.1 (iii) to obtain strict convexity of the robustified objective function, hence uniqueness of the optimal robust solution. Then the statement follows immediately from Corollary 4.1 together with Theorem 3.3. \square

The preceding theorem remains true in case $\lambda = 1$, if uniqueness of the maximum return portfolio is assumed and if the robust counterpart has a unique solution for $\lambda = 1$, which is for instance the case if U is a full n -dimensional ellipsoid, cf. Schöttle [33].

5. Conclusion and outlook. In this paper, we have shown that the robust counterpart to an already well-posed problem remains well-posed under some mild regularity and uniqueness assumption. As a consequence, we have established sufficient conditions under which the robust counterpart converges to the original solution, if the level of robustification is decreased. Based on the well-posedness of the robust counterpart, it was also shown how any consistent plug-in estimator can be supplemented by a corresponding consistent robust estimator based on a proper choice of the confidence set of the plug-in estimator. Finally, this consistency result was applied to the case of mean-variance portfolio optimization, which leads to a generalization of already known consistency results in this particular field.

As next steps, it would be interesting to investigate in more details the following two questions, which we leave for future research:

- How are robustified plug-in estimators related to the framework of robust estimation in the sense of Hampel and Huber, see for example Huber [16], where distributional uncertainty plays the major role?
- Is it possible to derive general statements on the (asymptotic) efficiency of the robust estimator in comparison to the original plug-in estimator?

For the second question, it would at least be necessary to establish asymptotic normality of the robustified plug-in estimators, which in turn would require an analysis of the smoothness properties of the robust counterpart with respect to the uncertainty set.

Acknowledgments. The author highly appreciates the valuable feedback by Katrin Schöttle on an early version of this paper.

References

- [1] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer. *Non-Linear Parametric Optimization*. Birkhäuser, Basel, 1983.
- [2] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, Princeton, 2009.
- [3] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.
- [4] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, 2nd edition, 1999.
- [5] P.-A. Bliman and C. Prieur. On existence of smooth solutions of parameter-dependent convex programming problems. In *Proceedings of 16th International Symposium on Mathematical Theory of Networks and Systems (MTNS2004)*. 2004.

- [6] J. F. Bonnans and A. Shapiro. Optimization Problems with Perturbations: A Guided Tour. *SIAM Rev.*, 40(2):228–264, 1998.
- [7] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, New York, 2000.
- [8] L. Breiman. *Probability*. SIAM, Philadelphia, 1992.
- [9] S. Ceria and R. A. Stubbs. Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management*, 7(2):109–127, 2006.
- [10] J. L. Doob. *Measure Theory*. Springer, New York, 1993.
- [11] K.-T. Fang, S. Kotz, and K.-W. Ng. *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London, 1990.
- [12] K.-T. Fang and Y.-T. Zhang. *Generalized multivariate analysis*. Science Press, Beijing, 1990.
- [13] D. Goldfarb and G. Iyengar. Robust Portfolio Selection Problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.
- [14] J. A. Gomez and W. Gomez. Cutting plane algorithms for robust conic convex optimization problems. *Optimization Methods and Software*, 21(5):779–803, 2006.
- [15] J. Hadamard. *Lectures on Cauchy’s Problem in Linear Partial Differential Equations*. Yale University Press, New Haven, 1923.
- [16] P.-J. Huber. *Robust Statistics*. John Wiley & Sons, New York, 1981.
- [17] J.-E. Ingersoll. *Theory of Financial Decision Making*. Rowman & Littlefield, Maryland, 1987.
- [18] J. D. Jobson and B. Korkie. Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, 75:544–554, 1981.
- [19] R. Kan and D. R. Smith. The distribution of the sample minimum-variance frontier. *Management Science*, 54(7):1364–1380, 2007.
- [20] R. Kan and G. Zhou. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656, 2007.
- [21] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*. Springer, Berlin, 1996.
- [22] D. G. Luenberger. *Optimization by vector space methods*. John Wiley & Sons, New York, 1969.
- [23] F. Lutgens. *Robust Portfolio Optimization*. PhD Thesis, Maastricht University, 2004.
- [24] H. Markowitz. Portfolio Selection. *Journal of Finance*, 7(1):77–91, 1952.
- [25] A. Meucci. *Risk and Asset Allocation*. Springer, Berlin, 2005.
- [26] H. Mori. Finite sample properties of estimators for the optimal portfolio weight. *J. Japan Statist. Soc.*, 34(1):27–46, 2004.
- [27] P. H. Müller. *Wahrscheinlichkeitsrechnung und Mathematische Statistik. Lexikon der Stochastik*. Akademie Verlag, Berlin, 5th edition, 1991.
- [28] Y. Okhrin and W. Schmid. Distributional Properties of Portfolio Weights. *Journal of Econometrics*, 134:235–256, 2006.
- [29] C. Perret-Gentil and M.-P. Victoria-Feser. Robust Mean-Variance Portfolio Selection. FAME Research Paper no 140, 2004.
- [30] S. J. Press. *Applied Multivariate Analysis*. Holt, Rinehart & Winston, New York, 1972.
- [31] H. Rinne. *Taschenbuch der Statistik*. Harri Deutsch, Frankfurt, 3rd edition, 2003.
- [32] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1997.
- [33] K. Schöttle. *Robust optimization with application in asset management*. Dissertation Technische Universität München, 2007.
- [34] K. Schöttle and R. Werner. Towards reliable efficient frontiers. *Journal of Asset Management*, 7(2):128–141, 2006.
- [35] K. Schöttle and R. Werner. Robustness properties of mean-variance portfolios. *Optimization*, 58(6):641–663, 2009.
- [36] R. H. Tütüncü and M. Koenig. Robust Asset Allocation. *Annals of Operations Research*, 132(1–4):157–187, 2004.
- [37] R. Werner. Cascading: an adjusted exchange method for robust conic programming. *Central European Journal of Operations Research*, 16(2):179–189, 2008.