Costs and benefits of robust optimization

Ralf Werner
Fakultät für Informatik und Mathematik
Hochschule München
email: werner@cs.hm.edu

First version: 11.11.2005 / This version: 31.10.2010

Abstract

In this exposition the robust counterpart approach by Ben-Tal, El Ghaoui and Nemirovski is investigated with respect to its costs and benefits, with the focus on the costs of robustification. Although robust optimization has gained more and more interest among both academics and practitioners and although this certainly represents a well-established theory, it is to some extent unclear, if and what costs have to be beared when using the robust counterpart formulation. Further, it is not known if other benefits besides the obvious can be realized by robustification. Concerning benefits, there is only one theoretical result by El Ghaoui and Lebret in the special situation of robust least squares besides the obvious robustness in feasibility under perturbations. In addition, on the cost side, one of the earlier papers [3] by Ben-Tal and Nemirovski provides a stability analysis together with a result concerning costs for robust linear optimization under convex uncertainty. In the following, Ben-Tal and Nemirovski's results on costs are generalized to smooth convex conic problems under Lipschitz uncertainty, given reasonably mild regularity conditions. For robust linear optimization, it is shown that under affine uncertainty and ellipsoidal uncertainty set, uniqueness of the optimal robust solution may be achieved as additional benefit in most situations.

1 Introduction

Over the last ten years the robust counterpart approach, which was introduced and made popular by Ben-Tal, Nemirovski [3, 4, 5] and independently by El Ghaoui et al. [7, 8], gained more and more interest among both academics and practitioners. An extensive introduction to the topic can be found in the recent book by Ben-Tal, Nemirovski and El Ghaoui [2], which also contains a large lists with references demonstrating the popularity of this approach. Although this methodology can be seen as well-established, a few questions remain open. For instance, it is to some extent unclear, if and what the benefits are – besides the obvious robustified feasibility – when moving from the original problem formulation to the robust counterpart. As only exception we note the interesting result on robust least squares by El Ghaoui and Lebret [7], Section 6. Further, it is not immediate how to quantify the costs of this robustification. Even though there is a tremendous amount of literature available dealing with robust optimization, most of this work either focuses on the numerical tractability or is concerned with a specific application in the gist of Ben-Tal and Nemirovski [5], but without general theoretical results.

To the best of our knowledge the only paper addressing the question of costs of robust optimization is one of the earlier papers [3], by Ben-Tal and Nemirovski themselves.

In [3], a stability analysis of robust optimization for linear problems under convex[1] uncertainty is provided. In this case, it is shown (see Proposition 2.3 in [3]) that the difference in the optimal values of the original (nominal) instance and the robust counterpart is of the order of the diameter of the uncertainty set, if a regularity condition (i.e. strict feasibility of the nominal LP) holds. Although no explicit statement about the convergence of the optimal solutions of the robust counterpart to a solution of the nominal instance for decreasing level of robustness (i.e. decreasing diameter of the uncertainty set) is 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. Besides Proposition 2.3, Ben-Tal and Nemirovski only provide one additional result on the costs of the robust counterpart: in their Theorem 2.1 it is shown that the optimal robust objective equals the worst-case objective of all *generalized nominal instances* under affine uncertainty. Unfortunately, due to the different understanding of the notion of a nominal instance by Ben-Tal and Nemirovski (which we call generalized instance in the following to avoid confusion), this Theorem 2.1 does not hold as expected but needs careful interpretation, as already pointed out by Ben-Tal and Nemirovski themselves, see [4], Section 2.2.

Subsequently, in Example 3.1 we will provide a counterexample to [3], Theorem 2.1 (similar to the one by Ben-Tal and Nemirovski [4], Section 2.2), based on the usual notion of an instance, which is then used to demonstrate that [3], Theorem 2.1 cannot be generalized to conic problems. In addition, this exposition will make the following main contributions to the topic of robust optimization:

- It will be shown in Theorem 3.4 that the result of [3], Proposition 2.3 can be generalized from linear problems with convex uncertainty to convex conic problems with arbitrary uncertainty under a similar regularity condition (a Slater condition) and global Lipschitz continuity. In this way, it will be established that robustification always comes at costs in the order of the diameter of the uncertainty set.
- Besides this asymptotic result, a stronger, non-asymptotic upper bound on the costs of robustification is provided in Theorem 3.5. This bound holds for all uncertainty sets beneath a certain diameter, where the diameter can be obtained from information about the Slater point of the nominal instance. The bounding constant will further be based on the (dual) solution of an auxiliary problem.
- It will be expatiated that the benefits (besides the obvious) of the robustification depend on the geometry of the uncertainty set, as well as on the way how uncertainty enters the problem. For the case of linear problems with affine uncertainty in the coefficients, it will be shown that polyhedral uncertainty sets do not effect the problem structure, whereas ellipsoidal uncertainty has a positive effect similar to Tikhonov regularization.

These findings come on top of some of our recent results concerning continuity properties of the robust counterpart, see Werner [13]. One of the main results in [13] is closely related to the results concerning costs here: it was shown that for decreasing level of robustness, both the optimal robust objective value and the optimal robust solution converge to those of the nominal instances, if some mild regularity condition is satisfied. However, no statement concerning the rate of this convergence was made.

The rest of the paper is organized as follows. After a succinct introduction to the topic, the local robust counterpart is introduced together with a similar problem, called

[1]Actually, Ben-Tal and Nemirovski work under the assumption of *concave* uncertainty, which means – due to their different notation of the constraint $g(x,u) \geq 0$ as opposed to our notation $g(x,u) \leq 0$ – convex uncertainty in our setup.

strong robust counterpart. The properties of these problems will be discussed in details in the second section. In Section 3, the main relations between the family of problem instances and the robust counterpart are established. Results concerning both the gap between the robust counterpart and the nominal instances and the costs of robustification of a nominal problem instance will be derived. As main result, it will be shown that under suitable regularity conditions, the costs decrease linearly with the size of the uncertainty. In Section 4 the benefits of robustification are considered in more detail. After arguing that no general result concerning benefits can be anticipated, a few basic observations on robust linear optimization will be made which allow to obtain uniqueness of the optimal robust solution in most cases.

2 The local robust counterpart

The following notation is used throughout the paper: $\mathcal{K}(M)$ denotes the space of all non-empty, convex and compact subsets of a given non-empty and convex set $M \subset \mathbb{R}^n$. Equipped with the usual Hausdorff distance, denoted by $d_H(A, B)$ for $A, B \in \mathcal{K}(M)$, this becomes a metric space. Further, $B_\varepsilon(x)$ denotes the closed ε-ball around some vector $x \in \mathbb{R}^n$ and $||x||$ denotes the 2-norm for $x \in \mathbb{R}^n$.

2.1 Setup and general assumptions

Throughout the paper let us consider the following finite dimensional convex 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 are identical to the assumptions made in the original paper [3] by Ben-Tal and Nemirovski:

- Let us assume throughout that $X \in \mathcal{K}(\mathbb{R}^n)$ and let $u \in \bar{U}$ with $\bar{U} \in \mathcal{K}(\mathbb{R}^d)$. Here, u denotes the uncertain data within a *global uncertainty set*[2] $\bar{U}$.
- Let $K \subset \mathbb{R}^m$ be a proper ordering cone with dual cone K^* and non-empty interior, i.e. $\text{int}(K) \neq \emptyset$. This means that parameter dependence is only modeled for inequality constraints.
- Let the objective $f : X \times \bar{U} \to \mathbb{R}$ and the constraint $g : X \times \bar{U} \to \mathbb{R}^m$ be continuously differentiable on $X \times \bar{U}$. We will see in the following that this smoothness assumption can be weakened to global Lipschitz continuity, i.e. there exists an $L_0 > 0$ such that

  $$\begin{aligned} |f(y, v) - f(x, u)| &\leq L_0 \left(||y - x|| + ||v - u||\right) \\ ||g(y, v) - g(x, u)|| &\leq L_0 \left(||y - x|| + ||v - u||\right) \end{aligned}$$

  for all $x, y \in X$ and $u, v \in \bar{U}$, cf. Proposition 2.1(vi).

[2] As noted in Ben-Tal and Nemirovski [3], $\bar{U}$ can be replaced by $\text{conv}(\bar{U})$ if g is K-convex in u for all x. However, as we will allow for a very general structure in u, we will assume convexity of the uncertainty set instead.

- 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 $u \in \bar{U}$.

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 $g(x,u) \in -K$. Please note that no structure is imposed on the way the uncertain parameter u enters the objective or constraint function.

Remark 2.1. In general, it is reasonable to assume that uncertainty enters in a convex way, as it is expected that the worst uncertain parameter for a given point x lies on the boundary of $\bar{U}$, which is the case for convex uncertainty. Taking into account the different constraint notation, convex uncertainty was also assumed in Ben-Tal and Nemirovski [3].

To avoid pathological situations in the investigations later on, we require that for all uncertain parameters u there exists at least one feasible point, i.e. for all $u \in \bar{U}$ there exists an $x(u) \in X$ with $g(x(u), u) \in -K$.

2.2 Local robust counterpart

Analogously to the definition of (P^*) in [3] where the whole uncertainty set $\bar{U}$ is used, let us define the *local robust counterpart* of a family of optimization problems $(P_u)_{u \in U}$ for some *local uncertainty set* $U \in \mathcal{K}(\bar{U})$ as

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

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 – which is the obvious benefit of robustification. It can be noted that in the special case $U = \{u\}$ the local robust counterpart coincides with the original problem instance (P_u).

For convenience of the proofs later on let us subsequently derive an equivalent convex reformulation of both (P_u) and (RC_U) with only one real-valued inequality constraint. Although this reformulation comes at the price of non-smoothness of the constraint, this allows for an easy analysis of the costs of both the original conic problem as well as the semi-infinite conic robust counterpart. As only global Lipschitz continuity will be necessary for this purpose, the loss of differentiability of the constraint g is not of great importance in the following.

2.3 Problem reformulation

For the purpose of reformulation[3], let us introduce the function $\tilde{g}_z : X \times \bar{U} \to \mathbb{R}$ for a given constraint g. The structure of $\tilde{g}_z$ is based on ideas in Gomez and Gomez [9] and Werner [12]:

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

where a positive anchor $z \in \operatorname{int}(K)$ is chosen arbitrarily, but fixed. In Gomez and Gomez [9] 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.

[3] For more details on this reformulation, let us refer to Werner [13], Section 2.2

Proposition 2.1. *Let the assumptions of Section 2.1 be satisfied and let L_0 be the global Lipschitz constant of g. Then it holds:*

(i) $\tilde{g}_z(x,u)$ *is continuous on* $X \times \bar{U}$ *and convex in* x *for all* $u \in \bar{U}$.

(ii) $\{x \in X \,|\, g(x,u) \in -K\} = \{x \in X \,|\, \tilde{g}_z(x,u) \leq 0\}$ *for all* $u \in \bar{U}$.

(iii) $\{x \in X \,|\, g(x,u) \in \operatorname{int}(-K)\} = \{x \in X \,|\, \tilde{g}_z(x,u) < 0\}$ *for all* $u \in \bar{U}$.

(iv) $\{x \in X \,|\, g(x,u) \in -K, \forall u \in U\} = \{x \in X \,|\, \tilde{g}_z(x,u) \leq 0, \forall u \in U\}$ *for all* $U \in \mathcal{K}(\bar{U})$.

(v) $\{x \in X \,|\, g(x,u) \in \operatorname{int}(-K), \forall u \in U\} = \{x \in X \,|\, \tilde{g}_z(x,u) < 0, \forall u \in U\}$ *for all* $U \in \mathcal{K}(\bar{U})$.

(vi) $\tilde{g}_z$ *is globally Lipschitz continuous on* $X \times \bar{U}$ *with Lipschitz constant* $L > 0$*:*

$$|\tilde{g}_z(y,v) - \tilde{g}_z(x,u)| \;\leq\; L(||y-x|| + ||v-u||)$$

for all $x, y \in X$ *and* $u, v \in \bar{U}$*, where* $L := L_0 \cdot \max\limits_{\substack{\lambda \in K^* \\ z^\top \lambda = 1}} ||\lambda||$.

(vii) If g *is* K*-convex in* u *for all* $x \in X$*, then* $\tilde{g}_z$ *is convex in* u *for all* $x \in X$.

Proof. Statements (i) to (v) are shown in Werner [13], Proposition 2.1. Statement (vii) follows from Rockafellar [11], Theorem 5.5. To see statement (vi) recall that

$$||g(y,v) - g(x,u)|| \;\leq\; L_0(||y-x|| + ||v-u||)$$

for all $x, y \in X$ and $u, v \in \bar{U}$ and therefore

$$\tilde{g}_z(y,v) = \max_{\substack{\lambda \in K^* \\ z^\top \lambda = 1}} \lambda^\top g(y,v) = \max_{\substack{\lambda \in K^* \\ z^\top \lambda = 1}} \lambda^\top g(x,u) + \lambda^T(g(y,v) - g(x,u)).$$

Then, as

$$-L_0||\lambda|| \cdot (||y-x|| + ||v-u||) \leq \lambda^T(g(y,v) - g(x,u)) \leq L_0||\lambda|| \cdot (||y-x|| + ||v-u||)$$

for all λ, the claim follows with $L := L_0 \cdot \max\limits_{\substack{\lambda \in K^* \\ z^\top \lambda = 1}} ||\lambda||$. □

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

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

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

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

which coincides with the robust counterpart of (P'_u). A better understanding of the reformulation may be gained from the following equality (see Werner [12] 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 [12], these distances lead to tractable formulations for LPs, SOCPs and SDPs, if the anchor is properly chosen.

2.4 Properties of the robustified problem

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

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

which are well-defined (i.e. finite everywhere). Then (RC'_U) becomes

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

The subsequent proposition collects the main properties of the robustified objective F and the robustified constraint G:

Proposition 2.2. *Under the assumptions of Section 2.1 it holds that:*

(i) F and G remain convex in x for each local uncertainty set U.

(ii) F and G are monotone in the uncertainty U, i.e. it holds

$$F(x,V) \leq F(x,U) \quad \text{and} \quad G(x,V) \leq G(x,U) \qquad \forall x \in X, \forall V \subset U.$$

This relationship especially holds for all $V = \{u\}$ with $u \in U$.

(iii) F and G are continuous on $X \times \mathcal{K}(\bar{U})$, i.e. it holds

$$F(x_n, U_n) \to F(x,U) \quad \text{and} \quad G(x_n, U_n) \to G(x,U)$$

for all $U_n \in \mathcal{K}(\bar{U})$ with $d_H(U_n, U) \to 0$ and $x_n \to x$.

(iv) F and G are globally Lipschitz continuous on $X \times \mathcal{K}(\bar{U})$:

$$\begin{aligned} |F(y,V) - F(x,U)| &\leq L(||y-x|| + d_H(U,V)) \quad \text{and} \\ |G(y,V) - G(x,U)| &\leq L(||y-x|| + d_H(U,V)) \end{aligned}$$

for all $x, y \in X$ and $U, V \in \mathcal{K}(\bar{U})$.

Proof. Statements (i) and (iii) can be found in Theorem 2.1 in Werner [13]. Statement (ii) is obvious. For the proof of statement (iv):

$$F(y,V) - F(x,U) = F(y,V) - F(y,U) + F(y,U) - F(x,U)$$

and therefore

$$|F(y,V) - F(x,U)| \leq |F(y,V) - F(y,U)| + |F(y,U) - F(x,U)|.$$

Now, as

$$\begin{aligned}
F(x,U) - F(y,U) &= \max_{u\in U} f(x,u) - \max_{u\in U} f(y,u) \\
&= \max_{u\in U} \big(f(x,u) - f(y,u) + f(y,u)\big) - \max_{u\in U} f(y,u) \\
&\leq \max_{u\in U} \big(L||y-x|| + f(y,u)\big) - \max_{u\in U} f(y,u) \\
&= L||y-x||
\end{aligned}$$

and vice versa (by swapping the roles of x and y in the above), we obtain

$$|F(y,U) - F(x,U)| \leq L||y-x||.$$

To get the second part of the claimed inequality, let $\mathbf{P}_U(v)$ denote the projection of v onto U. Then, as $\mathbf{P}_U(v) \in U$, it holds

$$\begin{aligned}
F(y,V) &= \max_{v\in V} f(y,v) = \max_{v\in V} \big(f(y,v) - f(y,\mathbf{P}_U(v)) + f(y,\mathbf{P}_U(v))\big) \\
&\leq \max_{v\in V} \big(L||v - \mathbf{P}_U(v)|| + f(y,\mathbf{P}_U(v))\big) \\
&\leq \max_{v\in V} L||v - \mathbf{P}_U(v)|| + \max_{v\in V} f(y,\mathbf{P}_U(v)) \\
&\leq Ld_H(U,V) + \max_{u\in U} f(y,u) \\
&= F(y,U) + Ld_H(U,V),
\end{aligned}$$

and swapping the roles of u and v proves the other direction. The same line of proof holds for the robust constraint G. □

2.5 Strong robust counterpart

In addition to the local robust counterpart, let us introduce the *strong robust counterpart* around u with robustness level t,

$$\begin{aligned}
\min_{x\in X} \quad & F(x,u,t) \\
\text{s.t.} \quad & G(x,u,t) \leq 0,
\end{aligned} \qquad (SRC_{u,t})$$

where

$$F(x,u,t) := f(x,u) + t, \quad \text{and} \quad G(x,u,t) := \tilde{g}_z(x,u) + t.$$

Using the notation

$$\begin{aligned}
\Phi(u) &:= \{x \in X \,|\, \tilde{g}_z(x,u) \leq 0\}, \\
\Phi(U) &:= \{x \in X \,|\, G(x,U) \leq 0\}, \\
\Phi(u,t) &:= \{x \in X \,|\, G(x,u,t) \leq 0\},
\end{aligned}$$

for the feasible sets with the obvious relation $\Phi(u) = \Phi(\{u\}) = \Phi(u,0)$, and similarly

$$\begin{aligned}
f^*(u) &:= \min_{x\in\Phi(u)} f(x,u), \\
F^*(U) &:= \min_{x\in\Phi(U)} F(x,U), \\
F^*(u,t) &:= \min_{x\in\Phi(u,t)} F(x,u,t),
\end{aligned}$$

for the optimal value functions with $f^*(u) = F^*(\{u\}) = F^*(u,0)$, we obtain the following

Lemma 2.1. *Let the assumptions from Section 2.1 be satisfied and let L be the joint global Lipschitz constant of f and $\tilde{g}_z$. Then it holds that*

$$\Phi(u, L\,\mathrm{diam}(U)) \subset \Phi(U) \subset \Phi(u)$$

and furthermore

$$F^*(u, L\,\mathrm{diam}(U)) \geq F^*(U) \geq f^*(u)$$

for all $U \in \mathcal{K}(\bar{U})$ and $u \in U$.

Proof. All statements directly follow from the global Lipschitz continuity, i.e. from

$$\tilde{g}_z(x,u) \leq G(x,U) \leq G(x,u,L\,\mathrm{diam}(U)) = \tilde{g}_z(x,u) + L\,\mathrm{diam}(U)$$

for all $x \in X$ and all $U \in \mathcal{K}(\bar{U})$ and $u \in U$. □

This immediately shows that the behavior of the local robust counterpart for small U (i.e. for $\mathrm{diam}(U) \to 0$) is intimately connected to continuity properties of $\Phi(u,t)$ for $t \to 0$, which in turn is closely related to a regularity condition at the optimal solution of (P_u). This connection is for instance used in the derivation of Theorems 3.3, Theorem 3.4 and Theorem 3.5.

3 Relations between (RC_U) and $(P_u)_{u\in U}$

In this section we want to establish the connection between the family of optimization problems $(P_u)_{u\in U}$ and their robust counterpart (RC_U). We start with some general observations and basic relations, similar to Ben-Tal and Nemirovski [3], which hold in a very general setup.

3.1 Basic relations

If the robust counterpart is feasible for some local uncertainty set U, then according to Lemma 2.1, we have $f^*(u) \leq F^*(U)$ for all $u \in U$, and therefore

$$\sup_{u\in U} f^*(u) \leq F^*(U).$$

For this property, we have only used the fact that the robustified functions F and G majorize the original functions f and $\tilde{g}_z$ for all $u \in U$. In this manner, we have proved the following

Theorem 3.1. *Let (RC_U) be feasible and let the assumptions from Section 2.1 be satisfied. Then all instances of (P_u) are feasible, and the optimal value $F^*(U)$ is greater than or equal to the optimal value $f^*(u)$ of all instances of (P_u).*

Note that Theorem 3.1 generalizes the result by Ben-Tal and Nemirovski (Proposition 2.2 in [3]) from convex uncertainty to general uncertainty and to general uncertainty sets U.

Remark 3.1. Let us point out that in Theorem 3.1 the meaning of an *instance* of problem (P_u) is the usual understanding of a problem instance, i.e. one uncertain parameter $u \in U$ is chosen and inserted in all functions (objective f and constraints g) at the same time.

For the reverse direction of Theorem 3.1, let us recall Theorem 2.1 from Ben-Tal and Nemirovski [3], which covers the special case $K = \mathbb{R}^m_+$ and affine uncertainty, i.e. the case

$$\begin{aligned} \min_{x \in X} \quad & f(x,u) && (Q_u) \\ \text{s.t.} \quad & g_1(x,u) \le 0, \\ & \vdots \\ & g_m(x,u) \le 0, \end{aligned}$$

where g_i are affine in u. Now, for Theorem 2.1 in [3] to hold, Ben-Tal and Nemirovski introduced the following *generalized instance* of problem (Q_u),

$$\begin{aligned} \min_{x \in X} \quad & f(x,u_0) && (Q_{u_0,u_1,\ldots,u_m}) \\ \text{s.t.} \quad & g_1(x,u_1) \le 0, \\ & \vdots \\ & g_m(x,u_m) \le 0, \end{aligned}$$

with an enlarged parameter space $U^{m+1} = U \times \cdots \times U$, i.e. a generalized instance is henceforth represented by different uncertain parameters for each individual constraint. In this case, Ben-Tal and Nemirovski's Theorem 2.1 states that if all instances of $(Q_{u_0,u_1,\ldots,u_m})$ are feasible, instead of only considering all instances of (P_u), then the robust counterpart is also feasible and the optimal values coincide. Clearly, in such a context, much more instances have to be considered than for the ordinary interpretation of a problem instance. Note that the reason for this generalization is naturally motivated by the fact that the robust counterparts of both problem families coincide, i.e. it does not matter if uncertainty is considered jointly or constraint-wise.

With these preliminaries, the proper reverse direction to Theorem 3.1 is the following

Theorem 3.2. *Let the assumptions from Section 2.1 hold. Then the robust counterpart of the problem family (Q_u) coincides with the robust counterpart of the generalized problem family $(Q_{u_0,u_1,\ldots,u_m})$. Further, in case of concave uncertainty, the robust counterpart is feasible if and only if all generalized instances $(Q_{u_0,u_1,\ldots,u_m})$ are feasible. In this case, the robust optimal value is the supremum of those of the generalized instances, i.e. it holds that*

$$F^*(U) = \sup_{(u_0,u_1,\ldots,u_m) \in U^{m+1}} f^*\big((u_0,u_1,\ldots,u_m)\big).$$

Proof. See Ben-Tal and Nemirovski [3], Theorem 2.1 for the case of affine uncertainty. The generalization from affine uncertainty to concave uncertainty is straightforward. □

Note that Theorem 3.2 does not state that the supremum will be attained, as supported by Example 3.1 below. Unfortunately, Theorem 3.2 cannot be extended to the conic case, not even under affine uncertainty[4], as the following Example 3.1 demonstrates. It also shows that the feasibility of the robust counterpart cannot be obtained from the feasibility of all usual instances (Q_u), not even for affine uncertainty, as already pointed out by Ben-Tal and Nemirovski themselves, see [4], Section 2.2.

[4]Shortly, if we started with a conic problem, even with affine uncertainty, after applying the reformulation trick from the previous section, we would obtain convex uncertainty instead of concave uncertainty as needed for Theorem 3.2.

Example 3.1. Consider the following convex problem under affine uncertainty:

$$\min_{x\in[-1;1]} \quad (1-u)\cdot(x-1)^2 \qquad (Q_u^1)$$
$$\text{s.t.} \quad ux \le 0$$

with $u \in \bar{U} = [-1;1]$. It can be immediately noted that there exists an optimal solution $x^*(u)$ for all $u \in \bar{U}$:

$$x^*(u) = \begin{cases} 1 & \text{if } u < 0 \\ 1 & \text{if } u = 0 \\ 0 & \text{if } u > 0 \end{cases} \quad \text{with} \quad f^*(u) = \begin{cases} 0 & \text{if } u < 0 \\ 0 & \text{if } u = 0 \\ 1-u & \text{if } u > 0. \end{cases}$$

We can equivalently rewrite (Q_u^1) as a problem where the uncertainty only enters the constraints, by introducing the auxiliary variable $y \in \mathbb{R}$:

$$\min_{\substack{x\in[-1;1]\\ y\in[-1;2]}} \quad y \qquad (Q_u^2)$$
$$\text{s.t.} \quad ux \le 0,$$
$$(1-u)\cdot(x-1)^2 - y \le 0.$$

Obviously, each instance of (Q_u^2) is feasible (by setting $y(u) := f^*(u)$) with the same optimal value as before. The worst optimal value is given by

$$\sup_{u\in[-1;1]} f^*(u) = 1,$$

however, there is no $u \in \bar{U}$ where this worst value is attained. Evidently, we can add the constraint $y \le 1$ without altering the problem and we obtain

$$\min_{\substack{x\in[-1;1]\\ y\in[-1;2]}} \quad y \qquad (Q_u^3)$$
$$\text{s.t.} \quad ux \le 0,$$
$$(1-u)\cdot(x-1)^2 - y \le 0,$$
$$y \le 1.$$

Now, choosing $u_1 = 1$ for the first inequality constraint and $u_2 = -1$ for the second inequality constraint under uncertainty shows that the corresponding generalized instance is not feasible, as the following system of inequalities does not have a feasible solution:

$$x \le 0,$$
$$2(x-1)^2 \le y,$$
$$y \le 1.$$

Therefore, according to Theorem 3.2, the robust counterpart of (Q_u^3) cannot be feasible, although each individual (usual) instance is feasible.

To illustrate that Theorem 3.2 cannot be extended to the conic case, let us rewrite problem (Q_u^3) as a conic problem by combining the three inequality constraints into one conic constraint

$$\min_{\substack{x\in[-1;1]\\ y\in[-1;2]}} \quad y \qquad (Q_u^4)$$
$$\text{s.t.} \quad \begin{pmatrix} ux \\ (1-u)\cdot(x-1)^2 - y \\ y-1 \end{pmatrix} \in -K_Q$$

with $K_Q = \mathbb{R}^3_+$ as corresponding cone. By this reformulation, the generalized instances are exactly the usual instances as there is only one conic constraint. Thus, each usual and each generalized instance is feasible, but the robust counterpart is not, which shows that Theorem 3.2 breaks down for conic constraints, even under affine uncertainty.

3.2 The gap between (RC_U) and $(P_u)_{u\in U}$

As seen above, in the general conic case, if only information about the feasibility of each (generalized) instance is at hand, not much can be said about the feasibility of the robust counterpart. Thus, we will subsequently assume that the robust counterpart is already feasible if not stated otherwise. For the following analysis of the gap between the robust counterpart (RC_U) and the worst-case of the (usual) instances $(P_u)_{u\in U}$ let us introduce both the *gap* Δ and the *relative gap* Δ_r for any $U \in \mathcal{K}(\bar{U})$ with $\operatorname{diam}(U) > 0$:

$$\begin{aligned} \Delta(U) &:= F^*(U) - \sup_{u\in U} f^*(u), \\ \Delta_r(U) &:= \frac{1}{\operatorname{diam}(U)}\Delta(U). \end{aligned}$$

As we have seen in Theorem 3.1, it always holds that $\Delta(U) \geq 0$. Further, as the robust counterpart of (Q^3_u) is not feasible we have $F^*(U) = +\infty$ and therefore $\Delta(U) = +\infty$ as well. We note that this can even happen in the case of affine constraint-wise uncertainty when Theorem 3.2 holds. However, one might still expect that $\Delta(U) \to 0$ if the uncertainty shrinks to a nominal instance, i.e. if $\operatorname{diam}(U) \to 0$, or equivalently $U \to \{u\}$ for some u. In this case, it would also be interesting to obtain results on the speed of convergence by investigating $\Delta_r(U) \to \text{const}$. Unfortunately, the robust counterpart of (Q^3_u) remains infeasible for all local uncertainty sets $U = [-\delta;\delta]$ and thus $\Delta(U) \not\to 0$ for $\operatorname{diam}(U) \to 0$ for a convex problem under affine uncertainty. This means that we at least have to assume that the local robust counterpart is feasible and that furthermore either a linear problem structure is given or some regularity condition is fulfilled. As the subsequent example demonstrates, the gap does not necessarily vanish with reduced uncertainty, not even in the case of a linear problem under affine uncertainty, and not even in case that the local robust counterpart is feasible. Therefore, it can be expected that the gap vanishes only if some regularity condition is satisfied for the original problem instances – and we will see that then the local robust counterpart automatically becomes feasible for sufficiently small uncertainty.

Example 3.2. For a fixed scaling parameter $s \in [1,\infty[$, consider the following linear problem under affine uncertainty:

$$\begin{aligned} \min_{x,y\in[-s;s]} \quad & x + y \\ \text{s.t.} \quad & ux \leq 0, \\ & -uy \leq 0. \end{aligned}$$

It can be immediately noted that each instance is feasible (let $x = y = 0$) and that the optimal solution and the optimal value are given by

$$\big(x^*(u), y^*(u)\big)^\top = \begin{cases} (0,-s)^\top & \text{if } u < 0 \\ (-s,-s)^\top & \text{if } u = 0 \\ (-s,0)^\top & \text{if } u > 0 \end{cases} \quad \text{with} \quad f^*(u) = \begin{cases} -s & \text{if } u < 0 \\ -2s & \text{if } u = 0 \\ -s & \text{if } u > 0. \end{cases}$$

Then, for an uncertainty set $U = [-\delta; \delta]$ with arbitrarily small $\delta > 0$, the local robust counterpart is given as

$$\begin{aligned} \min_{x,y \in [-s,s]} \quad & x + y \\ \text{s.t.} \quad & |x| \leq 0, \\ & |y| \leq 0, \end{aligned}$$

with obvious optimal solution $(x^*, y^*) = (0, 0)$ and optimal value 0. In this example, it holds that the gap can become arbitrarily large for a fixed uncertainty set U, depending on the scaling parameter:

$$\begin{aligned} \Delta(U) = \Delta([-\delta; \delta]) &= s \\ \Delta_r(U) = \Delta_r([-\delta; \delta]) &= \frac{1}{2\delta} s. \end{aligned}$$

More important, the gap remains constant independent of the size of U, i.e. it does not decrease to 0 for $U \to \{u\}$. This especially means that the relative gap becomes arbitrarily large, if the uncertainty is reduced to zero.

Now, consider the same example, but with different uncertainty set $U = [0; \delta]$. Then the robust counterpart becomes

$$\begin{aligned} \min_{x,y \in [-s;s]} \quad & x + y \\ \text{s.t.} \quad & x \leq 0, \\ & -y \leq 0. \end{aligned}$$

with optimal solution $(x^*, y^*) = (-s, 0)$ and optimal value $-s$. Thus, the gap equals 0 for all choices $\delta > 0$. However, even then, the robust counterpart does not converge to the nominal instance for $u = 0$.

3.3 The costs of robustification

As we have seen in Example 3.2, the gap Δ is not the right measure to describe the costs of robustification around a given nominal instance: the gap might vanish but there might remain a significant cost of robustification. Therefore, in addition to the gap function, we focus in the following on the *costs of robustification* of a nominal instance (P_u),

$$\begin{aligned} c(u, U) &:= F^*(U) - f^*(u), \\ c_r(u, U) &:= \frac{1}{\operatorname{diam}(U)} c(u, U), \end{aligned}$$

which can be be decomposed as

$$\begin{aligned} c(u, U) \;&=\; F^*(U) - f^*(u) \;=\; \big(F^*(U) - \sup_{v \in U} f^*(v)\big) + \big(\sup_{v \in U} f^*(v) - f^*(u)\big) \\ &=\; \Delta(U) + \big(\sup_{v \in U} f^*(v) - f^*(u)\big). \end{aligned} \tag{3.1}$$

From Equation (3.1) it becomes obvious that the costs of robustification can only decrease to 0, if the second term decreases to 0. However, this property is closely related to the continuity properties of the optimal value function f^* around the nominal instance:

Proposition 3.1. *Let the assumptions from Section 2.1 be satisfied. Then the mapping $\kappa : U \mapsto \sup_{v \in U} f^*(v) - f^*(u)$ is (Hausdorff) continuous in $U = \{u\}$ if and only if f^* is upper semicontinuous in u.*

Proof. As f^* is already lower semicontinuous based on the assumptions from Section 2.1, see Theorem 1.1 in Werner [13], the equivalence can be reduced to upper semicontinuity. If f^* is upper semicontinuous, then f^* is continuous and therefore κ is Hausdorff continuous according to Theorem 4.2.2(1) and (2) of Bank et al. [1], where now the set U takes the role of the parameter. For the reverse direction, if κ is Hausdorff continuous, then choose $U_n := \{u_n\}$ for any sequence $u_n \to u$, which proves the statement. □

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 for instance Bonnans and Shapiro [6], pp. 99). For example, according to Bank et al. [1], Corollary 4.4.1.3 or Bonnans and Shapiro [6], 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 $u = 0$.

Motivated by the above considerations, let us subsequently assume that the Slater condition holds:

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

Under the Slater condition it can be shown that the second term of the costs of robustification tends to 0 for decreasing level of robustness.

Proposition 3.2. *Let the assumptions from Section 2.1 be satisfied and let the Slater condition hold for (P_u). Then the mapping $\kappa : U \mapsto \sup_{v \in U} f^*(v) - f^*(u)$ is (Hausdorff) continuous in $U = \{u\}$ and it holds that*

$$\lim_{U_n \to \{u\}} \left(\sup_{v \in U_n} f^*(v) - f^*(u) \right) = 0.$$

Further, κ is locally Lipschitz continuous with some Lipschitz constant $\hat{L}$ in a neighborhood of $\{u\}$, i.e. there exists some $\delta > 0$ such that it holds

$$\kappa(U) \le \hat{L} \operatorname{diam}(U)$$

for all $U \in B_\delta(\{u\}) \subset \mathcal{K}(\bar{U})$, i.e. especially for all $U \in \mathcal{K}(\bar{U})$ with $u \in U$ and $\operatorname{diam}(U) \le \delta$.

Proof. Upper semicontinuity of f^* holds according to Bank et al. [1], Theorem 4.2.2, continuity then follows directly from Proposition 3.1. To see the local Lipschitz continuity, it is sufficient to show that f^* is locally Lipschitz continuous in u with Lipschitz constant $\hat{L}$, the rest then follows along the same lines as in the proof of Proposition 2.2(iv). Finally, local Lipschitz continuity of f^* follows from Theorem 4.25 in Bonnans and Shapiro [6], as both the objective f and the constraint g are assumed to be continuously differentiable. □

Examples 3.1 and 3.2 illustrate that both the gap between the supremum of the optimal value of all usual instances and the optimal value of the robust counterpart and the costs of robustification can become arbitrarily large. This can happen even if Theorem 3.2 holds. However, Theorem 3.4 below yields that the costs decrease linearly with the diameter of the uncertainty set, if the nominal instance satisfies the Slater condition, i.e. in this case both terms in Equation (3.1) shrink linearly to zero.

The following weaker result was proven in Werner [13], Theorem 2.3, where no global Lipschitz continuity of f and g was assumed and hence no linear decrease can be expected:

Theorem 3.3. *Let the assumptions from Section 2.1 be satisfied and let the Slater condition hold for* (P_u)*. Then the costs of robustification are (Hausdorff) continuous in* $U = \{u\}$ *and it holds that*

$$\lim_{U_n \to \{u\}} c(u, U_n) = 0.$$

In this case, for each sequence $U_n \to \{u\}$ *with corresponding optimal robust solution* $x_n^* := x^*(U_n)$*, each cluster point of the sequence* x_n^* *is optimal in* (P_u)*.*

For the much stronger result that the costs decrease at least linearly with the size of the uncertainty,

$$\limsup_{U_n \to \{u\}} c_r(u, U_n) = c_0 \in \mathbb{R}_+,$$

global Lipschitz continuity of all functions involved is needed as the following example shows:

Example 3.3. For the simple linear problem under non-Lipschitz uncertainty with $u \in [0; 1]$

$$\begin{aligned} \min_{x \in [0;2]} \quad & -x + 2\sqrt{u} \\ & x \leq 1 + \sqrt{u} \end{aligned}$$

it holds that $f^*(u) = \sqrt{u} - 1$ which is not Lipschitz continuous in $u = 0$. Further, for the local uncertainty set $U = [0; \delta]$ it holds that $F^*(U) = F^*([0; \delta]) = 2\sqrt{\delta} - 1$ and thus

$$\begin{aligned} \sup_{u \in [0;\delta]} f^*(u) &= \sqrt{\delta} - 1, \\ \Delta([0; \delta]) &= \sqrt{\delta}, \\ c(0, [0; \delta]) &= \sqrt{\delta} + \sqrt{\delta} = 2\sqrt{\delta}, \\ \Delta_r([0; \delta]) &= \frac{1}{\delta}\sqrt{\delta} = \frac{1}{\sqrt{\delta}}, \\ c_r(0, [0; \delta]) &= \frac{2}{\sqrt{\delta}}. \end{aligned}$$

This shows that even in such a simple case fulfilling all assumptions besides the global Lipschitz continuity, both the relative gap and the relative costs tend to infinity for $\delta \to 0$, although the gap and the costs themselves tend to 0 (in line with Theorem 3.3).

As we have assumed global Lipschitz continuity of all functions involved, the above example is not valid in our context, and the already announced result on asymptotically linear costs of robustification can be obtained.

Theorem 3.4. *Let the assumptions from Section 2.1 be satisfied and let problem* (P_u) *fulfill the Slater condition. Then it holds for all* $U \in \mathcal{K}(\bar{U})$ *with* $\delta := \operatorname{diam}(U)$ *sufficiently small (and* $u \in U$*) that*

$$f^*(u) \leq F^*(U) \leq F^*(B_\delta(u)) \leq f^*(u) + 2\mu_{\max} L\delta + o(\delta)$$

where $\mu_{\max}$ *is the largest Lagrange multiplier of the constraint* $\tilde{g}_z(x, u) \leq 0$ *in* (P_u)*. Furthermore, it holds that*

$$0 \leq \limsup_{U_n \to \{u\}} \Delta_r(U_n) \leq \limsup_{U_n \to \{u\}} c_r(u, U_n) \leq 2\mu_{\max} L,$$

i.e. both the gap and the costs decrease at least linearly with the order of the diameter of U *for small enough uncertainty sets.*

Proof. The first two inequalities are already known and only the last inequality

$$F^*(B_\delta(u)) \ \le f^*(u) + 2\mu_{\max} L\delta + o(\delta)$$

needs to be proved. Without loss of generality it can be assumed that the objective f does not depend on u, i.e. $f(x,u) = f(x)$. Otherwise, move the objective function to the constraint by introducing an auxiliary variable. Note that this move might enlarge the Lipschitz constant of the constraint by a factor of 2.

Then, as the nominal instance satisfies the Slater condition, it is obvious that the robust counterpart is feasible for δ sufficiently small. Now it holds that

$$G(x,U) \le G(x, B_\delta(u)) = \max_{\Delta u \in B_\delta(0)} \tilde{g}_z(x, u + \Delta u) \le \tilde{g}_z(x,u) + L\delta = G(x,u,L\delta)$$

according to Proposition 2.1(vi), where L is the global Lipschitz constant of $\tilde{g}_z$. Therefore, it holds that

$$\min_{x\in\Phi(u)} f(x) \ = \ f^*(u) \ \le \ F^*(U) \ \le \ F^*(B_\delta(u)) \ \le \ \min_{x\in\Phi(u,L\delta)} f(x)$$

where $\Phi(u, L\delta)$ represents the feasibility set of the strong robust counterpart with robustness level $L\delta$. This again shows the close relationship of the behavior of the local robust counterpart to the behavior of the feasible set $\Phi(u,t)$ for $t \to 0$. Let us consider this problem in more detail:

$$\min_{x\in\Phi(u,L\delta)} f(x) \quad = \quad \min_{\substack{x\in X \\ \tilde{g}_z(x,u)+t\le 0}} f(x) \ =: \varphi(t) \quad \text{at } \ t = L\delta.$$

As this problem coincides with (P_u) for $t = 0$, it also satisfies the Slater condition at $t = 0$. Therefore, we can apply Proposition 4.27 from Bonnans and Shapiro [6] and we obtain that φ is Hadamard directionally differentiable (in the direction $+1$) at $t = 0$ with Hadamard directional derivative

$$\varphi'(t) = \max_{\mu\in M(u)} \mu =: \mu_{\max}$$

where $M(u)$ denotes the set of all Lagrange multipliers of the constraint $\tilde{g}_z(x,u) \le 0$ in (P_u). Using the definition of the Hadamard directional derivative in $\mathbb{R}^1$, we obtain

$$\varphi'(t) = \lim_{t\to 0} \frac{\varphi(t) - \varphi(0)}{t} = \mu_{\max}$$

and therefore

$$\varphi(t) - \varphi(0) \le \mu_{\max} t + o(t)$$

for sufficiently small t. Setting $t = L\delta$ and using $F^*(U) - f^*(u) \le \varphi(L\delta) - \varphi(0)$ proves the statement. □

Although Theorem 3.4 already yields that the costs of robustification decrease at least linearly, this is still an asymptotic result. However, without much additional effort, the following result can also be established, which gives an upper bound on the cost of robustification:

Theorem 3.5. *Let the assumptions from Section 2.1 be satisfied and let problem (P_u) fulfill the Slater condition. Then, for $\delta := -\frac{1}{2}\tilde{g}_z(x_S(u), u)$ and $\gamma = \frac{\delta}{L}$,*

- *each local robust counterpart is feasible, and*

- $f^*(u) \le F^*(U) \le f^*(u) + 2L\mu \operatorname{diam}(U)$,

for all $U \in \mathcal{K}(\bar{U})$ *with* $u \in U$ *and* $\operatorname{diam}(U) \le \gamma$. *In this context,* μ *is some Lagrange multiplier of the constraint* $G(x, u, \delta) \le 0$ *in the strong robust counterpart. Under the given assumptions, the upper bound is sharp and cannot be improved any further.*

Proof. As in the proof of the previous theorem, we can assume again without loss of generality that the objective is not subject to uncertainty. Obviously, the strong robust counterpart with robustness level δ possesses the same Slater point as the original problem, as

$$G(x_S(u), u, \delta) = \tilde{g}_z(x_S(u), u) + \delta = \frac{1}{2}\tilde{g}_z(x_S(u), u) < 0.$$

Therefore, the local robust counterpart is feasible for $\operatorname{diam}(U) \le \gamma$ (and $u \in U$), as

$$G(x_S(u), U) \le \tilde{g}_z(x_S(u), u) + Ld_H(U, \{u\}) \le \\ \tilde{g}_z(x_S(u), u) + L \operatorname{diam}(U) \le \tilde{g}_z(x_S(u), u) + \delta.$$

Continuing with the strong robust counterpart instead of the original nominal instance, we define

$$\tilde{\varphi}(t) := \min_{\substack{x \in X \\ G(x,u,\delta) \le t}} f(x).$$

Note that t is now on the right hand side of the constraint, in contrast to the proof of the preceding theorem. This means that

$$\begin{aligned} \tilde{\varphi}(\delta) &= \varphi(0), \quad \text{and} \\ \tilde{\varphi}(0) &= \varphi(\delta), \end{aligned}$$

i.e. the optimal value of the original instance is given by $\tilde{\varphi}(\delta)$. Then, according to Luenberger [10], Theorem 1, p. 222:

$$\tilde{\varphi}(\delta) \ \ge \ \tilde{\varphi}(0) - \delta\mu,$$

or, equivalently

$$\varphi(\delta) \le \varphi(0) + \delta\mu,$$

for any Lagrange multiplier μ of the constraint $G(x, u, \delta) \le 0$ in the strong robust counterpart. With $F^*(U) \le \varphi(L \operatorname{diam}(U))$, $L \operatorname{diam}(U) \le \delta$ and $\varphi(0) = f^*(u)$, it follows that

$$f^*(u) \le F^*(U) \le f^*(u) + L \operatorname{diam}(U)\mu.$$

The sharpness of the bound is demonstrated by the following example. □

We note that the upper bound can be obtained by solving two optimization problems, first, the original nominal instance, and second, the strong robust counterpart with a robustness level determined by a suitable Slater point. The following example demonstrates that this bound is indeed sharp and cannot be improved under the given assumptions.

Example 3.4. Let us consider the linear problem

$$\min_{x \in [-1;1]} \quad x + u$$

$$-x + u \le 0$$

at $u = 0$. Obviously, all appearing functions are globally Lipschitz continuous with Lipschitz constant 1. Further, the optimal value of this problem is $f^*(u) = 2u$ and especially $f^*(0) = 0$. It is also easy to see that the local robust counterpart for $U = [-\delta; \delta]$ is given by

$$\begin{aligned} \min_{x\in[-1;1]} \quad & x + \delta \\ & -x + \delta \leq 0 \end{aligned}$$

with $F^*(U) = 2\delta$. Finally, some simple calculations show that the Lagrange multiplier to the constraint $-x + u \leq 0$ equals 1 for all u in some neighborhood of 0. Altogether, this example shows that $F^*(U) = 2\delta = 0 + 2 \cdot 1 \cdot \mathrm{diam}(U)$, i.e. the upper bound from Theorem 3.5 is attained and thus cannot be improved.

4 Benefits of robustification

In the following, a few results concerning the benefits of robustification will be provided for very specific situations. Unfortunately, not much can be said about the benefits of robustification in more general situations. Commonly, benefits could be categorized into the following three types:

- increased qualitative robustness,
- increased stability,
- increased quantitative robustness.

Besides the obvious gain in qualitative robustness by robustification, i.e. a robust feasible point is feasible for a whole range of uncertain parameters, one might hope that robustification gives rise to additional benefits. First of all, one might suppose that robustification would stabilize the optimal solution mapping, i.e. it is expected that the mapping $\hat{u} \mapsto x^*(U(\hat{u}))$ is continuous in the center of the uncertainty $\hat{u}$ in contrast to the – maybe discontinuous – mapping $u \mapsto x^*(u)$. Second, in case that the optimal solution mapping $u \mapsto x^*(u)$ is already Lipschitz continuous, one would anticipate a reduction in the corresponding Lipschitz constant. Unfortunately, both is not the case, as the following simple observation demonstrates.

Example 4.1. For the family of problem instances

$$\begin{aligned} \min_{x\in X} \quad & f(x) + u \\ & g(x) + u \leq 0 \end{aligned} \tag{R_u}$$

and the uncertainty sets $U(\hat{u}) = B_\delta(\hat{u}) \subset \mathbb{R}$, the corresponding robust counterparts are given by

$$\begin{aligned} \min_{x\in X} \quad & f(x) + \hat{u} + \delta \\ & g(x) + \hat{u} + \delta \leq 0. \end{aligned} \tag{$RCR_{U(\hat{u})}$}$$

Obviously, (assuming that everything is feasible) the robust problems are just shifted instances of the original nominal instances, i.e. $(R_{u+\delta}) = (RCR_{U(u)})$. Thus, no improvement in continuity or Lipschitz properties can be expected, as the optimal value and the optimal set mapping are just shifted by δ.

The above example clearly demonstrates that affine uncertainty has no significant impact on the problem structure. Apparently, to benefit from robustification, uncertainty has to enter objective and constraints in a much more direct fashion, of which the most simple form is the bilinear structure $u^\top x$. As we will see below, in this case, robustification actually leads to some improvements in the problem structure. Besides the following, in some other specific situations, it is also possible to derive positive results concerning the benefits of robustification. For instance, El Ghaoui and Lebret [7], Section 6, analyzed the application of the robust counterpart in the context of least squares. They were able to show, cf. Theorem 6.1 in [7], that the robust least squares solution is depending continuously on the problem data, and that the Lipschitz constant of this dependence decreases with increasing uncertainty. Of course, such an analysis always needs to be based on the very specific structure of the problem under consideration, and does only lead to a positive effect due to a favorable interaction of how the uncertainty set is chosen, how uncertainty enters the problem, and the problem structure itself.

Similarly, we will provide a slightly weaker result for problems with linear objective function under affine uncertainty in the coefficient, concerning the continuous dependence of the optimal solution on the uncertain data, cf. Theorem 4.1.

Proposition 4.1. *Under the assumptions from Section 2.1, let f be of the specific form*

$$f(x,u) = f_0(x) + (Au)^\top x$$

for some matrix $A \in \mathbb{R}^{n\times d}$ and some convex function f_0. Furthermore, let the local uncertainty set $U = U(\hat{u})$ have ellipsoidal shape, i.e.

$$U = \{\hat{u} + \delta H w \mid \|w\|_2 \leq 1\}$$

with $H \in \mathbb{R}^{d\times d}$ symmetric and positive semidefinite and $\delta > 0$. Then there exists an $x^* \in X$ such that one of the following holds for the set of optimal solutions

$$S(\hat{u}) := \{x \in \Phi(U(\hat{u})) \,|\, F(x, U(\hat{u})) \leq F^*(U(\hat{u}))\}$$

of the local robust counterpart:

(i) $S(\hat{u}) = \{x^*\}$, *i.e. the optimal solution is unique, or*

(ii) $S(\hat{u}) \subset \{y^* \mid y^* = \nu x^* + z, \nu \in \mathbb{R}, z \in \ker(HA^\top)\} \cap X$.

Proof. Easy calculations show that the robust objective function is given by

$$F(x, U(\hat{u})) = f_0(x) + (A\hat{u})^\top x + \delta ||HA^\top x|| \qquad (*)$$

which is again a convex function in x. Setting $f_1(x) := (A\hat{u})^\top x$ and $f_2(x) := \delta||HA^\top x||$ we see that each of the three functions f_0, f_1 and f_2 is itself convex in x. If x^* is the only optimal solution of the local robust counterpart, we are done; otherwise, let $y^* \neq x^*$ be a further optimal solution. Since the set of optimal solutions of a convex problem is convex, all points $\lambda x^* + (1-\lambda)y^*$ are optimal solutions as well with the same objective value, i.e.

$$F(x^*, U(\hat{u})) = F(y^*, U(\hat{u})) = F(\lambda x^* + (1-\lambda)y^*, U(\hat{u}))$$

for all $0 \leq \lambda \leq 1$. Further, due to convexity of f_0, f_1 and f_2

$$\begin{aligned}\lambda f_0(x^*) + (1-\lambda) f_0(y^*) &\geq f_0(\lambda x^* + (1-\lambda)y^*)\\ \lambda f_1(x^*) + (1-\lambda) f_1(y^*) &\geq f_1(\lambda x^* + (1-\lambda)y^*)\\ \lambda f_2(x^*) + (1-\lambda) f_2(y^*) &\geq f_2(\lambda x^* + (1-\lambda)y^*)\end{aligned}$$

it follows that equality holds in each row of the above inequality. This especially means that

$$||HA^\top(y^* + \lambda(x^* - y^*)|| = \lambda||HA^\top x^*|| + (1-\lambda)||HA^\top y^*||.$$

Since it is well-known that the 2-norm is strictly convex with only exception along half-rays starting from the origin, there exists an $\nu \in \mathbb{R}$ such that

$$HA^\top y^* = \nu HA^\top x^*$$

or equivalently

$$y^* = \nu x^* + z$$

where $z \in \ker(HA^\top)$, which proves the claim. □

Remark 4.1. The above result strongly depends on the shape of the uncertainty set to get rid of the linear behavior of the objective. If the uncertainty set is not assumed to be elliptical, but polyhedral (i.e. $U = \mathrm{conv}(u^1, \ldots, u^m)$), the statement of the theorem does not hold, as the following immediate observation shows,

$$\max_{u \in \mathrm{conv}(u^1,\ldots,u^m)} f_0(x) + (Au)^\top x \quad = \quad f_0(x) + \max_{1 \le k \le m} (Au^k)^\top x$$

which demonstrates that the robust objective is again a (piecewise) linear function.

Although this proposition does not seem to bring any immediate benefit for the stability of the optimal solution, it has some interesting consequences in case that every component of x is affected by the uncertainty. To guarantee that every component of x is affected by uncertainty, the uncertainty set is not allowed to be degenerate, i.e. it needs to be of the same dimension as x (i.e. $d = n$ and H is positive definite). Further, to ensure that every component of x is affected, the matrix A has to be regular.

Corollary 4.1. *Let the assumptions of Proposition 4.1 hold. If $d = n$, the matrix A is regular and H is positive definite, then the optimal solution set $S(\hat{u})$ of the local robust counterpart is either a singleton or a line segment:*

$$S(\hat{u}) \;=\; \{y^* \mid y^* = \nu\; x^*, \nu \in [\nu_l, \nu_u]\}$$

for some $\nu_l \le \nu_u \in \mathbb{R}$.

Proof. Straightforward, as $\ker(HA^\top) = \{0\}$ and X is compact. □

Theorem 4.1. *Let the assumptions of Corollary 4.1 hold and let $f(x, u) = c(u)^\top x$ with $c(u) = Au + c_0$. Then the optimal solution set $S(\hat{u})$ of the local robust counterpart is either a singleton or a line segment:*

$$S(\hat{u}) \;=\; \{y^* \mid y^* = \nu\; x^*, \nu \in [\nu_l, \nu_u]\}$$

for some $\nu_l \le \nu_u \in \mathbb{R}$. If further one of the following conditions hold

(i) $c_0 + A\hat{u} \in \mathbb{R}^n_+$ and $X \subset \mathbb{R}^n_+$,

(ii) $x \in X \implies \lambda x \notin X$ for all $\lambda \neq 1$,

then the solution of the local robust counterpart is unique. Furthermore, if problem (P_u) satisfies the Slater condition or if the feasible set $\Phi(u)$ is independent of the uncertain parameter u, then the optimal solution mapping $\hat{u} \mapsto x^*(U(\hat{u}))$ is continuous in a neighborhood of u.

Proof. The first statement follows from Corollary 4.1 with $f_0(x) = c_0^\top x$. Uniqueness of the optimal solution is immediate from (ii). To see uniqueness in case of (i), note that both $f_0(x) + f_1(x)$ and $f_2(x)$ in the robust objective (*) are positive, and that hence no multiple with larger norm can be part of the optimal solution. Continuity of the optimal solution mapping follows for example from Werner [13], Proposition 1.1. □

Interestingly, although looking quite artificial, condition (ii) is fulfilled in one of the most popular fields of application of robust optimization – robust portfolio optimization. Usually, in portfolio optimization, the asset allocation is modeled by the weights $x \in \mathbb{R}^n$ which usually have to add to 1, i.e. $\sum_{i=1}^n x_i = 1$, and therefore (ii) holds. Thus, the optimal solution to a robust portfolio allocation problem becomes unique, if such a *perpendicular constraint* is present in X, see Figure 4 for illustration. Together with

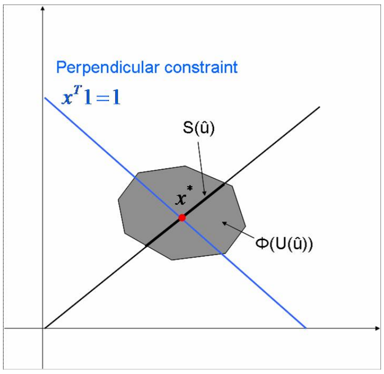

Figure 1: Illustration of the effect of adding a perpendicular constraint like $x^\top \mathbf{1} = 1$.

the existence of a Slater point for the original problem this means that – after robustification – the optimal solution is indeed continuously depending on the problem data, see e.g. Werner [13], Proposition 1.1. This means that the impact of robustification can be compared to the benefits which arise from a Tikhonov regularization of the objective function. This is not unexpected, as the specific structure of the robust objective function (see Equation (*)) indeed resembles a Tikhonov regularization, besides the missing square of the norm.

5 Summary

In this exposition we have shown that the costs of robustification usually decrease linearly with the size of the uncertainty set. For this purpose, we have generalized a result for linear problems under convex uncertainty by Ben-Tal and Nemirovksi to smooth convex problems under arbitrary (smooth) uncertainty. We have further demonstrated that the linear decrease does not only hold asymptotically, but is also valid in a whole neighborhood of the original instance. Based on a simple example it could be noted that this bound is indeed sharp and cannot be improved under the given quite weak assumptions. It remains open, if under stronger assumptions (e.g. a LICQ and a strong second order growth condition) some faster decline could be obtained or if some statements on the rate of the convergence of the optimal solutions can be made.

Concerning benefits, the specific situation of linear objective under affine uncertainty in the cost vector has been investigated. This leads to the insight that robustification is indeed beneficial in such a situation, as it induces uniqueness of the optimal solution – which in turn might result in a continuous dependence on the problem data. This is especially important in the context of portfolio optimization where robustification may often lead to better posed problems by inducing uniqueness and continuity of the optimal robust portfolio.

Acknowledgements
The author likes to thank Katrin Schöttle for valuable discussions. In addition, comments by participants of the Paraopt X conference on Theorem 3.5 are highly appreciated.

References

[1] Bank, B., Guddat, J., Klatte, D., Kummer, B., Tammer, K.: Non-Linear Parametric Optimization. Birkhäuser, Basel (1983)

[2] Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust Optimization. Princeton University Press, Princeton (2009)

[3] Ben-Tal, A., Nemirovski, A.: Robust convex optimization. Mathematics of Operations Research **23**(4), 769–805 (1998)

[4] Ben-Tal, A., Nemirovski, A.: Robust solutions of uncertain linear programs. Operations Research Letters **25**, 1–13 (1999)

[5] Ben-Tal, A., Nemirovski, A.: Robust Optimization – Methodology and Applications. Mathematical Programming **92**(3), 453–480 (2002)

[6] Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2000)

[7] El Ghaoui, L., Lebret, H.: Robust solutions to least-square problem with uncertain data matrices. SIAM Journal of Matrix Analysis and Application **18**, 1035–1064 (1997)

[8] El Ghaoui, L., Oustry, F., Lebret, H.: Robust solutions to uncertain semidefinite programs. SIAM Journal on Optimization **9**, 33–52 (1998)

[9] Gomez, J.A., Gomez, W.: Cutting plane algorithms for robust conic convex optimization problems. Optimization Methods and Software **21**(5), 779–803 (2006)

[10] Luenberger, D.G.: Optimization by vector space methods. John Wiley & Sons, New York (1969)

[11] Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1997)

[12] Werner, R.: Cascading: an adjusted exchange method for robust conic programming. Central European Journal of Operations Research **16**(2), 179–189 (2008)

[13] Werner, R.: Consistency of robust optimization with application to portfolio optimization (2010). www.optimization-online.org