

Reducing conservatism in Robust Optimization

Ernst Roos, Dick den Hertog

CentER and Department of Econometrics and Operations Research, Tilburg University

e.j.roos@tilburguniversity.edu, d.denhertog@tilburguniversity.edu

September 13, 2018

Abstract

Although Robust Optimization is a powerful technique in dealing with uncertainty in optimization, its solutions can be too conservative when it leads to an objective value much worse than the nominal solution or even to infeasibility of the robust problem. In practice, this can lead to robust solutions being disregarded in favor of the nominal solution. This conservatism is caused by both the constraint wise approach of Robust Optimization and its core assumption that all constraints are hard for all scenarios in the uncertainty set. This paper seeks to alleviate this conservatism by proposing an alternative robust formulation that condenses all uncertainty into a single constraint, binding the worst-case expected violation in the original constraints from above. Using recent results in distributionally robust optimization, the proposed formulation is shown to be tractable for both right- and left-hand side uncertainty. A computational study is performed with problems from the NETLIB library. For some problems, the percentage of uncertainty is magnified fourfold in terms of increase in objective value of the standard robust solution compared to the nominal solution, whereas we find solutions that safeguard against over half the violation while only a tenth of the uncertainty is reflected in the objective value. For problems with an infeasible standard robust counterpart, the suggested approach is still applicable and finds both solutions that safeguard against most of the uncertainty at a low price in terms of objective value.

Keywords: robust optimization, non-constraint wise uncertainty, ambiguity

1 Introduction

Most real-life optimization problems contain parameters that are not known precisely. Potential sources of this uncertainty are measurement, estimation and implementation errors in the underlying processes (Ben-Tal et al., 2009, p. xi). To deal with such uncertainty two classes of techniques exist: stochastic programming and robust optimization.

Stochastic programming is a class of techniques in which all uncertain parameters are assumed to follow a known probability distribution. Instead of regular (in)equalities, stochastic programming problems contain chance or expectation constraints. For a more detailed description we refer to the textbook by Shapiro et al. (2014). Unfortunately, stochastic programming has two major disadvantages. First of all, the exact distribution of uncertain parameters is often not known and thus must be estimated using historical data if available. This may lead to severe inaccuracy in the obtained solutions. Secondly, the problems that result from stochastic programming are generally hard to solve (Shapiro and Nemirovski, 2005).

Robust Optimization (RO), on the other hand, is a class of techniques that does not need any information on the distribution of the uncertain parameters and was initiated by Ben-Tal and Nemirovski (1998); El Ghaoui and Lebret (1997), although earlier papers on the topic exist, see for example, Soyster (1973). Instead, it requires the definition of an uncertainty set that contains all scenarios one wants to safeguard against. Subsequently, RO forces every constraint to be feasible for all possible parameter values in the uncertainty set. For an overview of RO and its applications we refer to Ben-Tal et al. (2009). An advantage of this approach is that the resulting problems are generally not much more difficult to solve than the original problem. For instance, techniques have been developed for nonlinear and even nonconvex RO (Mutapcic and Boyd, 2009; Bertsimas et al., 2010). From a more practical point of view, RO has been applied successfully to facility location problems (Baron et al., 2011; Gabrel et al., 2014) and network design problems (Mudchanatongsuk et al., 2008; Alvarez-Miranda et al., 2015; Pessoa and Poss, 2015) among many others.

Ben-Tal and Nemirovski (2000) showed that RO can successfully safeguard against potential constraint violation in multiple NETLIB problems. Out of 90 NETLIB problems studied, they found that 27 have a nominal solution that severely violate at least one constraint when considering a box uncertainty set that allows parameters to deviate by 1%. Some of the robust solutions they find, have an objective value up to 10% higher than the nominal objective value however, i.e., safeguarding against 1% of uncertainty is paid for tenfold in terms of objective value. This ‘price’ one pays for robustness has been addressed before, see e.g., Bertsimas and Sim (2004). Moreover, in extreme cases, this conservatism can lead to the standard robust counterpart being infeasible. In other words, robust optimization’s conservative nature might result in no solution that safeguards against uncertainty at all. Ben-Tal et al. (2017) mention this conservatism and discuss two approaches to deal with an infeasible robust counterpart: a simple approach that minimizes slacks similar to light robustness as introduced by Schöbel (2014) and an approach based on globalized robust optimization.

This trade-off is the main disadvantage of RO: it can lead to overly conservative solutions. This conservative nature of robust solutions has two major causes. First of all, one of the core assumptions of RO is that all constraints are ‘hard’ for all parameter values in the uncertainty set, which is often not the case in practice. Note that this assumption is twofold in the sense that it assumes that (1) no constraint can be violated for (2) any scenario in the uncertainty set. We remark that this assumption is relaxed by ‘light robustness’ (Fischetti and Monaci, 2009) and ‘globalized robust optimization’ (Ben-Tal et al., 2006, 2017). An overview of these (and other) methods is given by Goerigk and Schöbel (2016).

Light robustness addresses the conservatism of robust optimization by setting a limit to the deterioration of the objective value compared to the nominal solution. From all solutions that satisfy this limit, it then finds the solution that minimizes a weighted sum of all constraint violations. It is important to note that these violations are modeled using (non-adjustable) slack variables and thus this approach suffers from some of robust optimization’s conservatism through its constraint wise nature. This last observation is the main difference between light robustness and the approach we suggest.

Globalized robust optimization assumes that constraints are hard for all parameter values in a ‘normal range’ of scenarios and allows for violation in the constraints for scenarios that fall outside this ‘normal range’. It additionally assumes that the allowed violation for such a scenario is proportional to the distance from that scenario to the ‘normal range’. Globalized robust optimization thus requires specification of two uncertainty sets and treats them in a different way. Our approach, on the other hand, is limited to a single uncertainty set and allows a maximum amount of constraint violation for any perturbation in this set. Moreover, globalized robust optimization shares light robustness’ main difference to our approach: they do not account for the second cause of conservative solutions, the constraint wise nature of applying RO. Especially for uncertainty sets with high dependence between parameters spread over multiple constraints, this may lead to overly conservative solutions. The most obvious example is uncertainty on the right-hand side modeled with a budget uncertainty set (Bertsimas and Sim, 2004). RO will treat this uncertainty set exactly the same as box uncertainty for the right-hand side, that is, it completely ignores the budget constraint.

A more recent technique that deals with uncertainty in optimization combines ideas from stochastic and robust optimization and is referred to as Distributionally Robust Optimization (DRO). In general, it considers chance and/or expectation constraints that must hold for all probability distributions in some specified ambiguity set, that is, it considers not the worst-case but the worst-case expected behavior. Recent papers by Wiesemann et al. (2014) and Postek et al. (2018) have shown that problems with ambiguous

expectation constraints are computationally tractable in specific situations. Two existing approaches within the DRO framework that address the conservatism of robust optimization are ‘soft robust optimization’ (Ben-Tal et al., 2010) and ambiguous joint chance constraints, see e.g. (Chen et al., 2010).

Soft robust optimization is an approach using ambiguous expectation constraints that is similar to globalized robust optimization in the main idea. Where globalized robust optimization considers two distinct uncertainty sets, soft robust optimization considers an infinite nested family of ambiguity sets. More specifically, it considers nested sets of which the size is dependent on a parameter ϵ . Similar to globalized robust optimization, the allowed worst-case expected violation for distributions in a set with parameter ϵ is proportional to ϵ . The differences between our approach and soft robust optimization echo the differences with globalized robust optimization: it requires a thorough definition of a family of ambiguity sets and it does not account for the constraint wise nature of applying RO.

Alternatively, one can use ambiguous joint chance constraints to allow violation in constraints. In such an approach, one bounds the probability that violation occurs in any of the constraints from above by some prescribed number. By considering all original constraints in a single joint chance constraint, this approach circumvents the constraint wise nature of applying RO, similar to our approach. Problems involving chance constraints are, however, notoriously hard to solve, although improvements have been made recently (Chen et al., 2010). Moreover, different from the approach we propose, the amount by which constraints are violated is deemed irrelevant when using chance constraints.

In this paper, we present a general approach that allows violation of constraints to a certain extent and combines all constraints into one to find less conservative solutions. More specifically, we consider the total violation over all constraints and require it to be below a specified value. Choosing this parameter strictly positive then relaxes the assumption that all constraints are ‘hard’. Moreover, we show that under certain conditions the problem is equivalent to introducing slack variables for a budget uncertainty set. In general, however, it is not tractable for realistic problem sizes. For that reason, we focus on a different formulation that considers the expected violation instead. This approach is motivated by the recent developments in DRO mentioned above. Moreover, in problems that need to be solved more frequently, e.g., on a weekly basis, robustness might better be represented by the worst-case expected violation.

Our approach is tested on the NETLIB library for different amounts of constraint violation. For these problems, we are able to quantify and visualize the trade-off between objective value and constraint violation. For some problems, allowing for a 1% deviation

in the parameters can lead to an increase in objective value of over 4% in the standard robust solution. We, on the other hand, find solutions that safeguard against over half this uncertainty while only increasing in objective value by 0.1%.

The approach presented in this paper has the following advantages:

- It obtains solutions that safeguard against most uncertainty with a lower price of robustness than the standard robust solution.
- For problems with a non-unique optimal nominal solution, it is able to find an optimal nominal solution that also minimizes the worst-case expected violation.
- When the standard robust counterpart is infeasible, it finds the solution that minimizes the worst-case expected violation as well as solutions that safeguard against part of the uncertainty relatively cheaply.
- It is computationally tractable.

Section 2 outlines the general approach that is expanded on in the rest of the paper. Section 3 and 4 expand on the details of the approach for right- and left-hand side uncertainty, respectively. Section 5 presents numerical results based on problems from the NETLIB library and Section 6 concludes the paper.

2 Proposed General Approach

Consider a general linear optimization problem given by

$$\begin{aligned} \min_x \quad & c^\top x \\ \text{s.t.} \quad & Ax \geq b, \end{aligned} \tag{P1}$$

where $x \in \mathbb{R}^n$ is the decision vector, $c \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m \times n}$ are given parameters and $b \in \mathbb{R}^m$ are parameters that are uncertain. Throughout this paper we consider both uncertainty in A and b , but for now, we assume that all uncertainty considered is in b , the right-hand side of the constraints. The main reason for this assumption is that the conservatism mentioned in Section 1 is present in particular for right-hand side uncertainty. Intuitively, this is explained by the observation that the uncertain parameters are to some extent 'maximally' spread over a multitude of constraints and thus the constraint-wise nature of RO is most detrimental. We emphasize that everything stated in this section is also valid when A is considered to be the uncertain parameter. Moreover, we note that we can also treat the case where both A and b are uncertain, for example by introducing an extra variable x_{n+1} that is forced to equal 1 and models the right-hand side.

In RO, the uncertainty in b is modeled by letting it reside in a user-provided convex and compact uncertainty set U . It is then required that the constraints $Ax \geq b$ hold for all values $b \in U$, that is, the robust counterpart of the general linear optimization problem (P1) is

$$\begin{aligned} \min_x \quad & c^\top x \\ \text{s.t.} \quad & Ax \geq b \quad \forall b \in U. \end{aligned} \tag{P2}$$

To alleviate the constraint wise approach of RO, we combine all constraints of (P2) into a single constraint on the sum of all constraint violation. Moreover, we drop the assumption that all constraints are ‘hard’ and allow some constraint violation α . We also choose to enforce that the solution is feasible for the nominal or expected scenario \bar{b} to avoid solutions with excessive amounts of violation in a single constraint. The resulting formulation is given by

$$\begin{aligned} \min_x \quad & c^\top x \\ \text{s.t.} \quad & \sum_{i=1}^m \max \{0, b_i - a_i^\top x\} \leq \alpha \quad \forall b \in U \\ & Ax \geq \bar{b}. \end{aligned} \tag{P3}$$

Observe that for $\alpha = 0$, (P3) is equivalent to (P2). An important remark when considering (P3) is that by summing the violations, constraints for which b_i is larger in magnitude are automatically considered to be more important, as their effect on the total violation is larger. In applying this technique, it is therefore desirable to consider the relative importance of all constraints. Note that in many practical examples constraints will already be scaled accordingly. In a facility location problem for example, where an uncertain amount of demand should be satisfied, each unit of missed demand is generally of equal importance. If one does not know the exact interpretation of constraints, for example, the obvious approach would be to normalize each constraint such that $|\bar{b}_i| = 1$.

A disadvantage of this problem formulation is that the standard way to solve the robust optimization problem (P3) is only tractable for box and budget uncertainty sets on right-hand side parameters using a reformulation introduced by Ardestani-Jaafari and Delage (2016). This tractability issue stems from the fact that this robust counterpart contains the maximum over a convex function. Traditionally, (P3) could be reformulated

by introducing auxiliary variables to yield

$$\begin{aligned}
& \min_x c^\top x \\
& \text{s.t.} \quad \sum_{i=1}^m y_i \leq \alpha \\
& \quad y_i \geq b_i - a_i^\top x \quad i = 1, \dots, m, \quad \forall b \in U \quad (\text{P4}) \\
& \quad y_i \geq 0 \quad i = 1, \dots, m \\
& \quad Ax \geq \bar{b}.
\end{aligned}$$

For robust optimization problems, however, this equivalence is in general not true (Gorissen and Den Hertog, 2013). It is true, however, when U is a box uncertainty set or when variables y_i are considered to be adjustable variables. In the latter case, Ardestani-Jaafari and Delage (2016) show that for a budget uncertainty set, affine decision rules suffice, as this yields a problem equivalent to their reformulation. Unfortunately, even though this reformulation is a linear optimization problem, its number of constraints and variables grow quadratically in the number of uncertain parameters of the original problem. This means that in practice it can lead to impractical problem sizes.

One of the causes of RO's conservative solutions is that it only considers the worst case over the uncertainty set, while in reality there is not always a high probability of this worst case occurring. Recent papers in DRO therefore consider the uncertain parameters to follow an unknown distribution that resides in some ambiguity set. One can then consider the worst-case expected violation, that is, compute the expected violation with respect to the distribution in this ambiguity set for which it is highest. Recent papers in DRO show that this worst-case expectation can be computed efficiently for specific ambiguity sets. We thus consider the problem

$$\begin{aligned}
& \min_x c^\top x \\
& \text{s.t.} \quad \mathbb{E}_{\mathbb{P}} \left[\sum_{i=1}^m \max \{0, b_i - a_i^\top x\} \right] \leq \alpha \quad \forall \mathbb{P} \in \mathcal{P} \quad (\text{P5}) \\
& \quad Ax \geq \bar{b},
\end{aligned}$$

where \mathcal{P} is the ambiguity set that contains all distributions one considers for b . Although (P5) may seem to be a more difficult problem to solve than (P3), this is not necessarily true. Section 3 discusses an ambiguity set for which computing the worst-case expectation is remarkably efficient. Section 4 discusses this formulation in detail when A is the uncertain parameter. Note that the optimal value function of (P5) is convex in α .

Lemma 1. *The optimal value function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ of (P3) given by*

$$f(\alpha) = \min_x \left\{ c^\top x \mid \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\sum_{i=1}^m \max \{0, b_i - a_i^\top x\} \right] \leq \alpha, \quad Ax \geq \bar{b} \right\},$$

is convex.

Proof. We know the supremum as well as the sum of convex functions are convex functions itself (Rockafellar, 1997). The result then follows by Corollary 2.7 in Fiacco and Kyparisis (1986). \square

We note that given some objective value β , we can use all techniques described in Sections 3 and 4 to find the solution that minimize the worst-case expected violation with this objective value. Mathematically, this means we solve

$$\begin{aligned} \min_x \quad & \alpha \\ \text{s.t.} \quad & \mathbb{E}_{\mathbb{P}} \left[\sum_{i=1}^m \max \{0, b_i - a_i^\top x\} \right] \leq \alpha \quad \forall \mathbb{P} \in \mathcal{P} \\ & c^\top x \leq \beta \\ & Ax \geq \bar{b}. \end{aligned} \tag{P6}$$

This formulation can, for example, be used to find a nominal solution that minimizes the worst-case expected violation if the nominal solution is not unique. Moreover, if we omit the constraint on the original objective value ($c^\top x \leq \beta$), we can find the solution that minimizes the worst-case expected violation.

3 Right-Hand Side Uncertainty

The technique we consider to solve (P5) is the distributionally robust optimization approach developed by Postek et al. (2018). This technique assumes the ambiguity set to contain all probability distributions for b with a given support, mean and mean absolute deviation from the mean (MAD), for which all b_i are pairwise independent. We remark that the function of interest, the sum of maxima, is separable in the uncertain parameters, as for right-hand side uncertainty each constraint only contains a single uncertain parameter. Therefore, we do not require an assumption regarding their pairwise independence. To see this, note that the expectation is a linear operator and we can thus consider the maximum over the ambiguity set per individual maximum, i.e., only the marginal distributions of b_i , $i = 1, \dots, L$, matter. Mathematically, this means that we consider

$$\mathcal{P} = \{ \mathbb{P} : \text{supp}(b_i) \subseteq [l_i, u_i], \quad \mathbb{E}_{\mathbb{P}}(b_i) = \mu_i, \quad \mathbb{E}_{\mathbb{P}}|b_i - \mu_i| = d_i, \quad \forall i \}.$$

The required support, mean and MAD can often be estimated from historical data, which makes this approach suitable for practical applications. Moreover, if no information regarding the MAD is available, the approach we propose can still be applied. In this case, all techniques in this section can be used with $d_i = \frac{2(u_i - \mu_i)(\mu_i - l_i)}{u_i - l_i}$, which is the highest possible MAD any distribution with the specified support and mean can have. This yields the expressions as first derived by Madansky (1959).

Given that the ambiguity set is defined as above, we can reformulate (P5) using the following equalities by using (14) from Postek et al. (2018). In this derivation we use that the sum of maxima is a separable function as described above:

$$\begin{aligned} \max_{\mathbb{P} \in \mathcal{P}} \mathbb{E} \left[\sum_{i=1}^m \max \{0, b_i - a_i^\top x\} \right] &= \max_{\mathbb{P} \in \mathcal{P}} \sum_{i=1}^m \mathbb{E} [\max \{0, b_i - a_i^\top x\}] \\ &= \sum_{i=1}^m \max_{\mathbb{P} \in \mathcal{P}} \mathbb{E} [\max \{0, b_i - a_i^\top x\}] \\ &= \sum_{i=1}^m \sum_{y \in \{1,2,3\}} p_y^i \max \{0, \tau_y^i - a_i^\top x\}, \end{aligned} \quad (1)$$

where $\tau_1^i = l_i$, $\tau_2^i = \mu_i$ and $\tau_3^i = u_i$ are the support of the worst-case distribution and their probabilities are given by

$$p_1^i = \frac{d_i}{2(\mu_i - l_i)}, \quad p_2^i = 1 - \frac{d_i}{2(\mu_i - l_i)} - \frac{d_i}{2(u_i - \mu_i)}, \quad p_3^i = \frac{d_i}{2(u_i - \mu_i)},$$

for $i = 1, \dots, m$. The derivation above states that the worst-case expected violation constraint can simply be replaced by the latter expression resulting in the following problem

$$\begin{aligned} \min_x \quad & c^\top x \\ \text{s.t.} \quad & \sum_{i=1}^m \left[\sum_{y \in \{1,2,3\}} p_y^i \max \{0, \tau_y^i - a_i^\top x\} \right] \leq \beta. \end{aligned} \quad (\text{P7})$$

In other words, the worst-case distribution is a three-point distribution on the support and mean. (P7) can be easily transformed into a linear optimization problem with $3m$ extra variables with linearization techniques for maxima. The resulting linear optimization problem is then:

$$\begin{aligned} \min_{x, z} \quad & c^\top x \\ \text{s.t.} \quad & \sum_{i=1}^m \sum_{y \in \{1,2,3\}} p_y^i z_y^i \leq \beta \\ & z_y^i \geq 0 \quad i = 1, \dots, m, \quad y \in \{1, 2, 3\} \\ & z_y^i \geq \tau_y^i - a_i^\top x \quad i = 1, \dots, m, \quad y \in \{1, 2, 3\} \\ & Ax \geq \bar{b}. \end{aligned} \quad (\text{P8})$$

The most obvious parameter choice for the ambiguity set is such that $\mu_i = \bar{b}_i$. Forcing each constraint to be nominal feasible ($Ax \geq \bar{b}$) then means that $\tau_y^i - a_i^\top x \leq 0$ for $i = 1, \dots, m$, $y \in \{1, 2\}$, as $\tau_1^i \leq \tau_2^i = \bar{b}_i$. The problem thus reduces to containing only m auxiliary variables modeling the violation of each original constraint for $b_i = \tau_3^i$. Although this resembles a box uncertainty set, (P8) is only equivalent to (P4), for an appropriate choice of d_i and u_i . In particular, one needs to choose u_i equal to the highest value b_i can take in the uncertainty set used in (P4) and d_i such that $p_1^i = p_1^j$ for all i, j .

Postek et al. (2018) also describe how the best-case distribution can be found based additional information on $\beta_i = \mathbb{P}(b_i \geq \mu_i)$. This yields a two-point distribution with which the best-case expected violation can be computed. Combined with the worst-case distribution, this information can be used to find an upper bound on the cost of not knowing the uncertain parameter's actual distribution.

4 Left-Hand Side Uncertainty

4.1 General Approach

For left-hand side uncertainty, the standard approach using the technique described in Section 3 would lead to an optimization problem involving an exponential number of terms, $m \cdot 3^n$ to be exact. To alleviate this computational burden, we use the adaptation proposed in Postek et al. (2018) that deals with aggregate uncertainty. More specifically, we consider $a_{ij} = \bar{a}_{ij} (1 + \epsilon \zeta_{ij})$, that is, to be a nominal value times some perturbation. A more general dependence on ζ_{ij} can be assumed by allowing ϵ to vary, that is, use some ϵ_{ij} . The effects of this alteration will be expanded upon at the end of this section. We then define the ambiguity set by

$$\mathcal{P} = \{\mathbb{P} : \text{supp}(\zeta_{ij}) \subseteq [-1, 1], \mathbb{E}(\zeta_{ij}) = 0, \mathbb{E}|\zeta_{ij}| = d_{ij} \text{ for } i = 1, \dots, m, \quad j = 1, \dots, n\}.$$

Note that this definition does not require ζ_{ij} to be pairwise independent. Moreover, this definition of the ambiguity set implicitly assumes that the uncertainty is symmetric around the nominal value. To reduce the dimensionality issues of $n \cdot m$ uncertain parameters, we combine the uncertainty from all ζ_{ij} for a given i , such that we in fact consider the random variables $y_i(x, \zeta) = (\bar{a}_i \circ x)^\top \zeta_i$, where $\bar{a}_i \circ x$ denotes the element-wise multiplication, also called Hadamard product, of \bar{a}_i and x . We are thus interested in solving the following problem, using the approach outlined in Section 3 applied to the uncertain factors $y_i(x)$:

$$\begin{aligned}
& \min_x c^\top x \\
& \text{s.t.} \quad \sum_{i=1}^m \mathbb{E}_{\mathbb{P}} [\max \{b_i - \bar{a}_i^\top x + y_i(x, \zeta), 0\}] \leq \alpha \quad \forall \mathbb{P} \in \mathcal{P} \quad (\text{P9}) \\
& \quad \bar{A}x \geq b,
\end{aligned}$$

where $\bar{A} \in \mathbb{R}^{m \times n}$ is the matrix consisting of the elements \bar{a}_{ij} . Since $y_i(x, \zeta)$ depends linearly on ζ_{ij} , it is easy to find its support and mean, given by

$$\text{supp}(y_i(x, \zeta)) = [-\|\bar{a}_i \circ x\|_1, \|\bar{a}_i \circ x\|_1], \quad \mathbb{E}[y_i(x, \zeta)] = (\bar{a}_i \circ x)^\top \mathbb{E}[\zeta_i] = 0.$$

Its mean absolute deviation, however, cannot be found exactly. The following theorem outlines how bounds on the mean absolute deviation can help bound the worst-case expected violation as well, analog to Propositions 3 and 7 in Postek et al. (2018).

Theorem 1. *For any $\delta_i(x)$ such that $\mathbb{E}[|y_i(x, \zeta)|] \leq \delta_i(x)$, the worst-case expected violation can be bounded from above by*

$$\begin{aligned}
& \sum_{i=1}^m \left[\frac{\delta_i(x)}{2\|\bar{a}_i \circ x\|_1} \max \{0, 2(b_i - \bar{a}_i^\top x), b_i - \bar{a}_i^\top x + \epsilon\|\bar{a}_i \circ x\|_1\} \right. \\
& \quad \left. + \left(1 - \frac{\delta_i(x)}{\|\bar{a}_i \circ x\|_1}\right) \max \{0, b_i - \bar{a}_i^\top x\} \right].
\end{aligned}$$

Proof. First of all, Postek et al. (2018) show that the worst-case expectation is a non-decreasing function of the MAD. Therefore, using an upper bound $\delta_i(x) \geq \mathbb{E}[|y_i(x, \zeta)|]$ instead of the MAD will give an upper bound for the expectation in (P9). We can thus use this upper bound together with the worst-case three-point distribution introduced by Postek et al. (2018) to find:

$$\begin{aligned}
& \sum_{i=1}^m \max_{\mathbb{P}_i \in \mathcal{P}_i} \mathbb{E}[\max \{b_i - y_i(x, \zeta), 0\}] \\
& \leq \sum_{i=1}^m \left[\frac{\delta_i(x)}{2(\mathbb{E}[y_i(x, \zeta)] - \|\bar{a}_i \circ x\|_1)} \max \{0, b_i - \bar{a}_i^\top x - \epsilon\|\bar{a}_i \circ x\|_1\} \right. \\
& \quad + \frac{\delta_i(x)}{2(\|\bar{a}_i \circ x\|_1 - \mathbb{E}[y_i(x, \zeta)])} \max \{0, b_i - \bar{a}_i^\top x + \epsilon\|\bar{a}_i \circ x\|_1\} \\
& \quad + \left(1 - \frac{\delta_i(x)}{2(\mathbb{E}[y_i(x, \zeta)] - \|\bar{a}_i \circ x\|_1)} \right. \\
& \quad \quad \left. \left. - \frac{\delta_i(x)}{2(\|\bar{a}_i \circ x\|_1 - \mathbb{E}[y_i(x, \zeta)])}\right) \max \{0, b_i - \bar{a}_i^\top x\} \right]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^m \left[\frac{\delta_i(x)}{2\|\bar{a}_i \circ x\|_1} \max \{0, b_i - \bar{a}_i^\top x - \epsilon \|\bar{a}_i \circ x\|_1\} \right. \\
&\quad + \frac{\delta_i(x)}{2\|\bar{a}_i \circ x\|_1} \max \{0, b_i - \bar{a}_i^\top x + \epsilon \|\bar{a}_i \circ x\|_1\} \\
&\quad \left. + \left(1 - \frac{\delta_i(x)}{2\|\bar{a}_i \circ x\|_1} - \frac{\delta_i(x)}{2\|\bar{a}_i \circ x\|_1} \right) \max \{0, b_i - \bar{a}_i^\top x\} \right]. \tag{2}
\end{aligned}$$

Subsequently we use the fact that $\mathbb{E}\zeta_i = 0$ to simplify (2). Moreover, since $\|\bar{a}_i \circ x\|_1 \geq 0$, we find that

$$\max\{0, b_i - \bar{a}_i^\top x + \epsilon \|\bar{a}_i \circ x\|_1\} = 0 \Rightarrow \max \{0, b_i - \bar{a}_i^\top x - \epsilon \|\bar{a}_i \circ x\|_1\} = 0.$$

Conversely it thus holds that

$$\begin{aligned}
&\max \{0, b_i - \bar{a}_i^\top x - \epsilon \|\bar{a}_i \circ x\|_1\} = b_i - \bar{a}_i^\top x - \epsilon \|\bar{a}_i \circ x\|_1 \\
&\Rightarrow \max \{0, b_i - \bar{a}_i^\top x + \epsilon \|\bar{a}_i \circ x\|_1\} = b_i - \bar{a}_i^\top x + \epsilon \|\bar{a}_i \circ x\|_1.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\max \{0, b_i - \bar{a}_i^\top x + \epsilon \|\bar{a}_i \circ x\|_1\} + \max \{0, b_i - \bar{a}_i^\top x - \epsilon \|\bar{a}_i \circ x\|_1\} \\
&= \max \{0, 2(b_i - \bar{a}_i^\top x), b_i - \bar{a}_i^\top x + \epsilon \|\bar{a}_i \circ x\|_1\}.
\end{aligned}$$

This results in us finding (2) is equal to

$$\begin{aligned}
&\sum_{i=1}^m \left[\frac{\delta_i(x)}{2\|\bar{a}_i \circ x\|_1} \max \{0, b_i - \bar{a}_i^\top x - \epsilon \|\bar{a}_i \circ x\|_1\} \right. \\
&\quad + \frac{\delta_i(x)}{2\|\bar{a}_i \circ x\|_1} \max \{0, b_i - \bar{a}_i^\top x + \epsilon \|\bar{a}_i \circ x\|_1\} \\
&\quad \left. + \left(1 - \frac{\delta_i(x)}{\|\bar{a}_i \circ x\|_1} \right) \max \{0, b_i - \bar{a}_i^\top x\} \right] \\
&= \sum_{i=1}^m \left[\frac{\delta_i(x)}{2\|\bar{a}_i \circ x\|_1} \max \{0, 2(b_i - \bar{a}_i^\top x), b_i - \bar{a}_i^\top x + \epsilon \|\bar{a}_i \circ x\|_1\} \right. \\
&\quad \left. + \left(1 - \frac{\delta_i(x)}{\|\bar{a}_i \circ x\|_1} \right) \max \{0, b_i - \bar{a}_i^\top x\} \right]. \tag{3}
\end{aligned}$$

In other words, the worst-case distribution for $y_i(x, \zeta)$ is a three point distribution on its mean and extremes of its support, where the probability density on the extremes is proportional to the mean absolute deviation, which is bounded from above. \square

The results in this section would hardly change if one assumes $a_{ij} = \bar{a}_{ij}(1 + \epsilon_{ij}\zeta_{ij})$, that is, one assumes ϵ to vary. More specifically, one would then define aggregate random variables $y_i(x, \zeta) = (\bar{a}_i \circ \epsilon_i \circ x)^\top \zeta_i$, where $\epsilon \in \mathbb{R}^n$ is the vector consisting of all ϵ_{ij} for some i . This implies that the support of $y_i(x, \zeta)$ is given by

$$\text{supp}(y_i(x, \zeta)) = [-\|\bar{a}_i \circ \epsilon_i \circ x\|_1, \|\bar{a}_i \circ \epsilon_i \circ x\|_1],$$

and the expression in Theorem 1 changes accordingly.

4.2 Bounds for the MAD

The simplest way to find an upper bound as suggested by Postek et al. (2018) is the following:

$$\begin{aligned}
\mathbb{E} |y_i(x, \zeta) - 0| &= \mathbb{E} \left| (\bar{a}_i \circ x)^\top \zeta_i \right| = \mathbb{E} \left| \sum_{j=1}^n \bar{a}_{ij} x_j \zeta_{ij} \right| \\
&\leq \mathbb{E} \sum_{j=1}^n |\bar{a}_{ij} x_j \zeta_{ij}| = \sum_{j=1}^n \mathbb{E} |\bar{a}_{ij} x_j \zeta_{ij}| = \sum_{j=1}^n \mathbb{E} |\bar{a}_{ij} x_j| |\zeta_{ij}| \\
&= \sum_{j=1}^n |\bar{a}_{ij} x_j| \mathbb{E} |\zeta_{ij}| = \sum_{j=1}^n |\bar{a}_{ij} x_j| d_{ij} \leq \|\bar{a}_i \circ x\|_1 \cdot \max_j d_{ij}. \tag{4}
\end{aligned}$$

The quality of this bound and the resulting formulation will be discussed at the end of this section.

This bound is particularly useful as it greatly simplifies (3). More specifically, we can show the following simplification analog to Proposition 7 in Postek et al. (2018).

Theorem 2. *For $\delta_i(x) = \|\bar{a}_i \circ x\|_1 \cdot \max_j d_{ij}$, the optimal value of (P9) is bounded from above by the optimal value of the following linear optimization problem:*

$$\begin{aligned}
&\min_{w, x, y, z} c^\top x \\
&s.t. \quad \sum_{i=1}^m \frac{1}{2} \max_j d_{ij} w_i + (1 - \max_j d_{ij}) y_i \leq \alpha \\
&\quad w_i \geq 2 (b_i - \bar{a}_i^\top x) \quad i = 1, \dots, m \\
&\quad w_i \geq b_i - \bar{a}_i^\top x + \epsilon \sum_{j=1}^n z_{ij} \quad i = 1, \dots, m \\
&\quad z_{ij} \geq \bar{a}_{ij} x_j \quad i = 1, \dots, m, \quad j = 1, \dots, n \quad (P10) \\
&\quad z_{ij} \geq -\bar{a}_{ij} x_j \quad i = 1, \dots, m, \quad j = 1, \dots, n \\
&\quad y_i \geq b_i - \bar{a}_i^\top x \quad i = 1, \dots, m \\
&\quad y_i, w_i \geq 0 \quad i = 1, \dots, m \\
&\quad \bar{A}x \geq b.
\end{aligned}$$

Proof. Using $\delta_i(x) = \|\bar{a}_i \circ x\|_1 \cdot \max_j d_{ij}$ we find

$$\begin{aligned}
&\sum_{i=1}^m \max_{\mathbb{P}_i \in \mathcal{P}_i} \mathbb{E} [\max \{b_i - \bar{a}_i^\top x - y_i(x, \zeta), 0\}] \\
&\leq \sum_{i=1}^m \left[\frac{\delta_i(x)}{2 \|\bar{a}_i \circ x\|_1} \max \{0, 2 (b_i - \bar{a}_i^\top x), b_i - \bar{a}_i^\top x + \epsilon \|\bar{a}_i \circ x\|_1\} \right. \\
&\quad \left. + \left(1 - \frac{\delta_i(x)}{\|\bar{a}_i \circ x\|_1} \right) \max \{0, b_i - \bar{a}_i^\top x\} \right]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^m \left[\frac{\|\bar{a}_i \circ x\|_1 \cdot \max_j d_{ij}}{2\|\bar{a}_i \circ x\|_1} \max \{0, 2(b_i - \bar{a}_i^\top x), b_i - \bar{a}_i^\top x + \epsilon\|\bar{a}_i \circ x\|_1\} \right. \\
&\quad \left. + \left(1 - \frac{\|\bar{a}_i \circ x\|_1 \cdot \max_j d_{ij}}{\|\bar{a}_i \circ x\|_1}\right) \max \{0, b_i - \bar{a}_i^\top x\} \right] \\
&= \sum_{i=1}^m \left[\frac{1}{2} \max_j d_{ij} \cdot \max \{0, 2(b_i - \bar{a}_i^\top x), b_i - \bar{a}_i^\top x + \epsilon\|\bar{a}_i \circ x\|_1\} \right. \\
&\quad \left. + \left(1 - \max_j d_{ij}\right) \cdot \max \{0, b_i - \bar{a}_i^\top x\} \right]. \tag{5}
\end{aligned}$$

Introducing auxiliary variables, the optimal value of (P9) can then be bounded from above by the optimal value of (P10). \square

Much like for right-hand side uncertainty, the inclusion of the constraint $\bar{A}x \geq b$ means that for the appropriate choices for d_{ij} , (P10) is equivalent to the ordinary slack problem (P4) with a unit box uncertainty set on ζ_{ij} . This equivalence holds when $d_{ij} = d$ for all i, j to be precise.

Furthermore, if all uncertain parameters in the same constraint have equal mean absolute deviation, we can show that (P9) is equivalent to the derived upper bound (P10).

Theorem 3. *Given the ambiguity set*

$$\mathcal{P} = \{\mathbb{P} : \text{supp}(\zeta_{ij}) \subseteq [-1, 1], \mathbb{E}(\zeta_{ij}) = 0, \mathbb{E}|\zeta_{ij}| = d_i \text{ for } i = 1, \dots, m, j = 1, \dots, n\}, \tag{6}$$

(P9) is equivalent to (P10).

Proof. Consider the probability distribution $\mathbb{P} \in \mathcal{P}$ such that

$$\mathbb{P}(\zeta_i = \mathbf{0}) = 1 - d_i, \quad \mathbb{P}(\zeta_i = \text{sign}(a_i \circ x)) = \mathbb{P}(\zeta_i = -\text{sign}(a_i \circ x)) = \frac{d_i}{2} \quad \forall i,$$

such that we find $\mathbb{E}_{\mathbb{P}}[y_i(x, \zeta)] = 0$ and

$$\mathbb{E}_{\mathbb{P}}[|y_i(x, \zeta)|] = d_i \|a_i \circ x\|_1.$$

This shows there exists a distribution that attains the bound for the mean absolute deviation found in (4). Therefore, this distribution is the worst-case distribution in this ambiguity set and thus (P9) is equivalent to (P10). \square

Theorem 3 shows that the bound on the MAD of $y_i(x, \zeta)$ given in (4) is tight when no assumption on the (in)dependence of ζ_{ij} is made and $d_{ij} = d_i$ for all j . If we do assume pair-wise independence of ζ_{ij} the looseness of the bound is highly dependent on how $\max_j d_{ij}$ compares to other d_{ij} for a given i .

We note that the results from this section can also be used when the MAD is unknown. Similar to the approach for right-hand side uncertainty we then instead use the maximum possible MAD: $d_{ij} = 1$ for all i, j . In particular this also means that the bound we obtain is exact, since all uncertain parameters have equal mean absolute deviation.

Besides the simple bound given above, Postek et al. (2018) suggest three other bounds. Unfortunately, all these bounds are either not applicable to the problem we discuss or do not offer advantages over the bound discussed above in terms of tractability or quality. More specifically, they cannot be used to obtain a convex constraint or their convex reformulation is a worse bound than the one discussed above.

5 Numerical Results: NETLIB Problems

5.1 NETLIB Problems

To analyze the practical applicability of all the approaches considered here, we consider problems from the NETLIB library, similar to Ben-Tal and Nemirovski (2000). In total, 80 problems from this library were considered to exhibit uncertainty. Once again, we consider uncertainty in the right-hand side of the inequality constraints first in Section 5.2. Uncertainty in the left-hand side is discussed in Section 5.3.

To measure the performance of our method, we attempt to quantify the trade-off between objective value and constraint violation that is represented by the solutions it

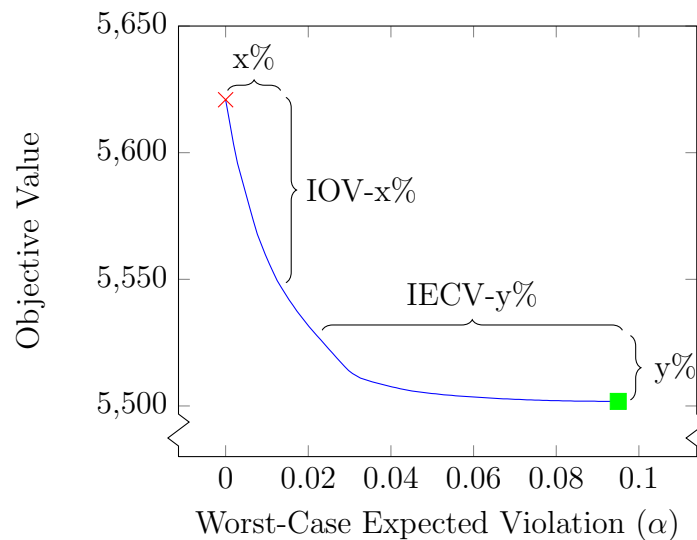


Figure 1: A graphical illustration of the IOV and IECV. The red cross indicates the standard robust solution and the green square indicates the nominal solution, while the blue line indicates the solutions found by our approach.

finds. More specifically, we report the potential improvement in violation compared to the nominal solution when a (slightly) higher objective value is allowed. Similarly, we report the potential improvement in objective value compared to the robust solution when some constraint violation is allowed. These measures will be denoted by Improvement Objective Value (IOV) and Improvement Expected Constraint Violation (IECV), followed by a percentage that indicates the allowed room in the constraint violation and objective value, respectively. Figure 1 illustrates these concepts graphically. In this figure $x\%$ is with respect to the nominal worst-case violation, while $y\%$ is calculated with respect to the nominal objective value. The IOV and IECV are also always reported in percentages with respect to the nominal solution.

It is important to note that the curve presented in Figure 1 is a Pareto curve in the sense that there exist no solutions that are nominal feasible with both a lower objective value and worst-case expected violation than any solution on the line. Moreover, we remark that this Pareto optimality of our solutions does not correspond to the Pareto robustly optimal solutions as defined by Iancu and Trichakis (2014). Their method can be applied to our problem, however.

Throughout this paper we assume parameters to deviate 1% from their nominal value, while Ben-Tal and Nemirovski (2000) also consider 0.1% and 0.01%. After running a representative selection of problems for other levels of uncertainty, we found there to be a simple effect on the results: multiplying the uncertainty by 10 yields the same tenfold increase in IOV- $x\%$ for any x . The IECV also adjusts accordingly, but since our chosen percentages by which to increase the objective value are suited to 1%, it is not particularly informative for other levels of uncertainty.

All numerical results have been obtained using MATLAB and Gurobi on a Lenovo Y700 with an i7-6700HQ and 16GB RAM.

5.2 Right-Hand Side Uncertainty

For right-hand side uncertainty, we only consider a parameter to be uncertain if it is different from zero, as a right-hand side of zero is often known with certainty. After selecting problems that do in fact have at least a single uncertain parameter by these standards, 64 problems remain. Of these 64 problems, there are 3 problems for which the nominal solution attains no violation whatsoever for all levels of uncertainty, due to the inequalities concerning the uncertain parameters not being binding in the nominal solution. These problems are therefore excluded from this analysis. This means we are left with 61 problems on which we have tested all techniques considered in this thesis that deal with right-hand side uncertainty.

As noted in Section 2, it is important to make sure all constraints are in the same order

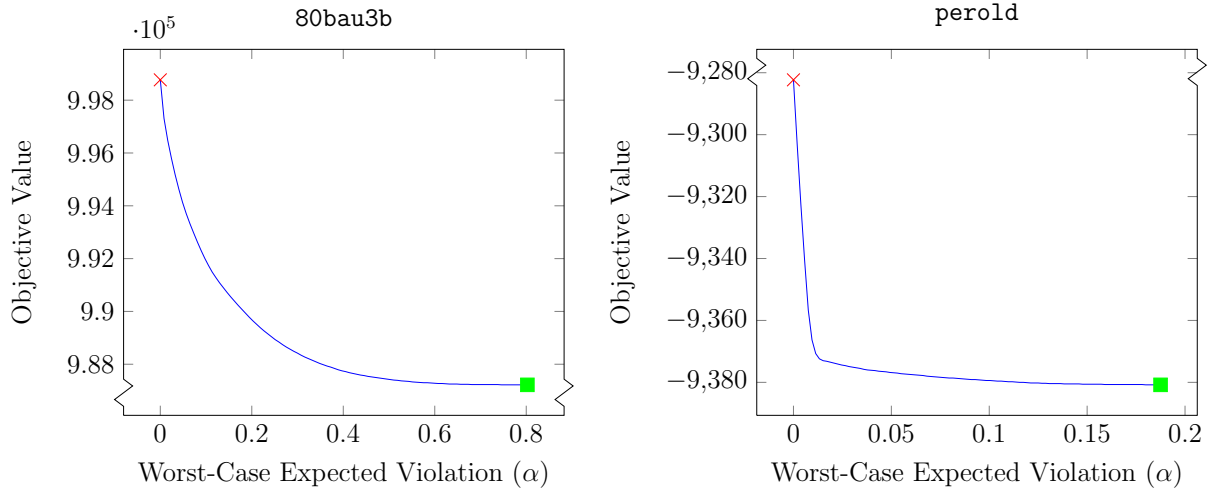


Figure 2: The solutions found by our approach with right-hand side uncertainty for 80bau3b and `sctap1`, shown in blue. The red cross indicates the standard robust solution and the green square indicates the nominal solution.

of magnitude such that their violation is comparable. Since we have no information on any constraints that are more important than others, we choose to normalize all constraints such that $|b_i| = 1$.

We consider the ambiguity set discussed in Section 3 where each uncertain parameter can deviate a fraction ϵ from its nominal value and has a mean absolute deviation proportional to this maximal deviation:

$$\mathcal{P} = \left\{ \mathbb{P} : \text{supp}(b_i) = [\bar{b}_i - \epsilon |\bar{b}_i|, \bar{b}_i + \epsilon |\bar{b}_i|], \mathbb{E}_{\mathbb{P}}(b_i) = \bar{b}_i, \mathbb{E}_{\mathbb{P}} |b_i - \bar{b}_i| = \delta \epsilon |\bar{b}_i| \right\},$$

where we specifically consider $\epsilon = 0.01$ and $\delta = 0.5$. The results of this approach are reported in Table 3 in the Appendix. As there is hardly any computational burden for this method, all solutions presented were obtained within seconds.

For five problems (`boeing`, `boeing2`, `degen2`, `degen3` and `ffff800`), the standard robust counterpart is infeasible for $\epsilon = 0.01$. For such a problem, the IOV cannot be computed. The ICV, however, is presented along with all other problems' characteristics in Table 3. First of all, note that the difference between the nominal and robust solution in terms of objective value is indicated in the table by IOV-100%. In other words, this number is the factor by which the uncertainty (here 1%) is magnified in the objective value for the standard robust solution.

A first observation of the table shows that for many problems, the IOV-50% and the IOV-100% are remarkably close. In other words, reasoning from the nominal solution, the first half of the worst-case constraint violation is usually very cheap to prevent, while the second half comes at a much higher price in terms of objective value. Moreover, the last

column shows that, except for **degen2**, **degen3** and **wood1p**, accepting an increase of 1% in objective value safeguards one against over half the worst-case constraint violation for all problems with 1% uncertainty. If regarded more closely, in fact a staggering 47 out of 61 problems have an IECV-1% higher than 90%, that is, for 47 problems over 90% of the worst-case expected constraint violation can be avoided by accepting an increase of 1% in objective value.

Another important remark is the fact that although any IECV that equals 100% indicates one can fully safeguard against the considered uncertainty, this is not necessarily the merit of our approach. More specifically, a 100% improvement in worst-case constraint violation simply indicates that the standard robust solution's objective value is within the set percentage from the nominal solution. The primary merit of our approach is found at problems such as **80bau3b** and **perold**, whose solutions are plotted in Figure 2. For **perold**, for example, we find that by safeguarding against only 90%, instead of all, of the possible worst-case violation, one can gain over 1% in objective value compared to the standard robust solution.

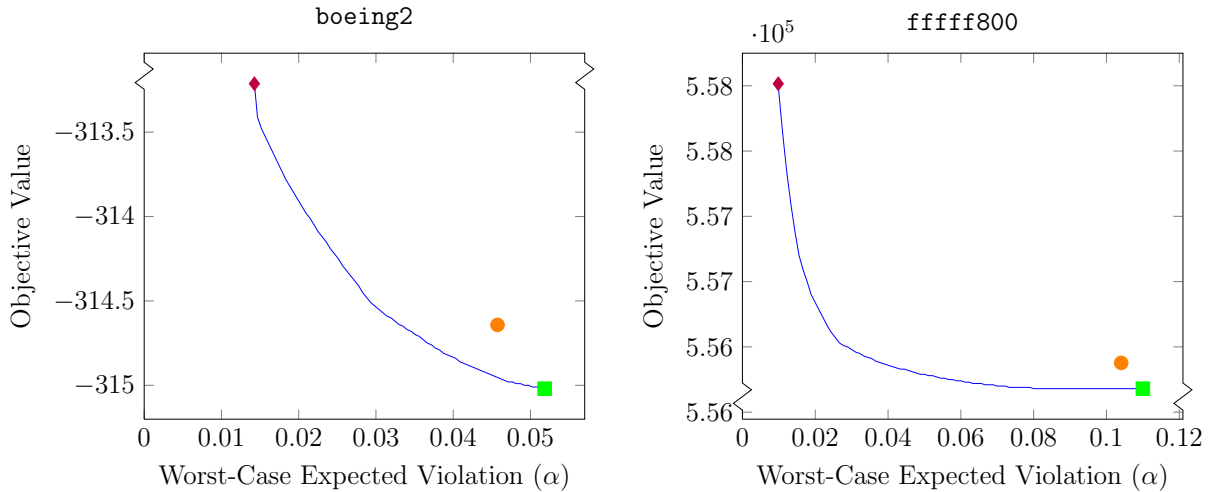


Figure 3: The solutions found by our approach with right-hand side uncertainty outlined for **boeing2** shown in blue. The green square indicates the nominal solution while the orange circle represents the solution found by shrinking the uncertainty set until we obtain feasibility for the standard robust counterpart. The purple diamond shows the solution found by minimizing the sum of slacks.

Figure 3 shows the solutions found for **boeing2** and **fffff800**, for which the standard robust counterpart is infeasible. The figure shows that, while we cannot find any solutions with a worst-case expected constraint violation equal to 0, the trade-off between violation and objective value can still be visualized for violation higher than that. Moreover, this approach allows one to find the solution with the least amount of worst-case expected

constraint violation possible under this uncertainty set. An intuitive way of dealing with an infeasible standard robust counterpart might be shrinking the uncertainty set (here by lowering ϵ) until it is feasible and using the solution found. The solution that is obtained in this way is shown as an orange circle in Figure 3. Clearly, this is not the best approach in dealing with an infeasible standard robust counterpart.

Another approach that can be used to deal with an infeasible standard robust counterpart is to introduce slack variables like in (P4) in all constraints, minimizing their sum and subsequently minimizing the objective value. The solution found by this approach is shown as a purple diamond in Figure 3. Note that, since we choose the mean absolute deviation such that it is proportional to the maximal deviation of a parameter and we assume nominal feasibility, this slack approach is equivalent to our approach and thus this solution lies on the end of the curve of solutions we find.

It is important to mention that the choice for δ has no effect whatsoever on the results, besides scale. This is due to the fact that δ only influences the probabilities of the three scenarios occurring and, since we enforce nominal feasibility here, thus simply scales the worst-case expected violation. Section 5.4 discusses relaxing the assumption that solutions should be nominal feasible.

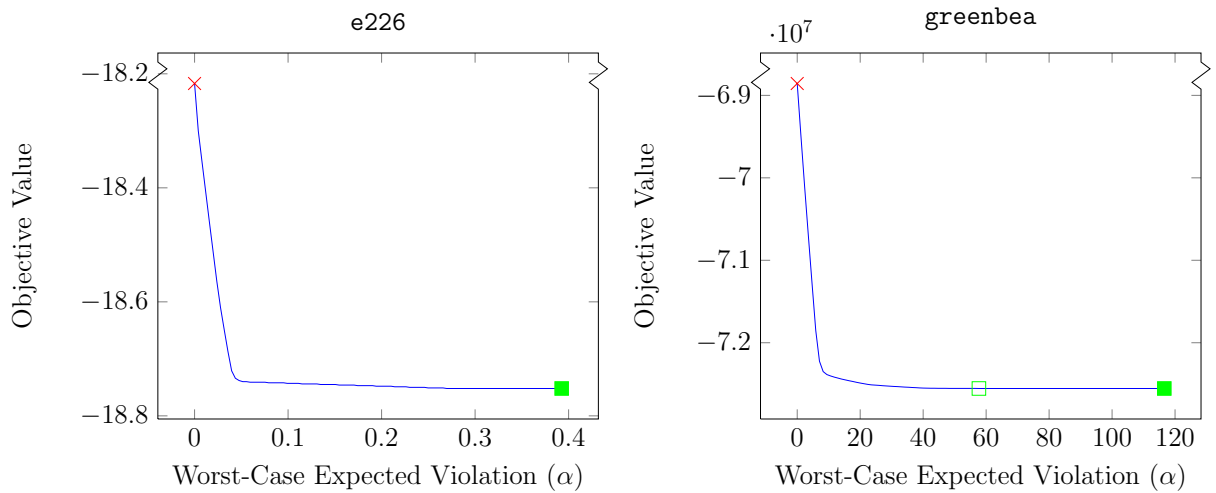


Figure 4: The solutions found by our approach with left-hand side uncertainty for **e226** and **greenbea**, shown in blue. The red cross indicates the standard robust solution and the filled green square indicates the nominal solution. For **greenbea**, the open green square indicates the most robust nominal solution.

5.3 Left-Hand Side Uncertainty

Similarly to Ben-Tal and Nemirovski (2000), we consider all entries of A that are specified to more than two decimals to be uncertain. By these standards, 49 of the 80 considered problems exhibit uncertainty. For three of these problems, the nominal solution has a worst-case expected violation of 0, which leaves 46 problems to analyze. Table 4 in the Appendix summarizes the results. We take (6) as the ambiguity set and thus assume equal MAD, i.e., $d_{ij} = d_i \forall j$. As discussed in Section 4.2 this means the IECV can be computed exactly.

Figure 4 shows the characteristics of the solutions found for **e226** and **greenbea**, for which the approach yields exceptionally good solutions. Both problems allow one to immunize against over 80% of the uncertainty at a mere 0.1% increase of the objective value. The standard robust solution, on the other hand, has an objective value 2.9% and 5.1% higher than the nominal solution, respectively.

For **greenbea**, the nominal solution is non-unique. In Figure 4 we have also indicated an optimal nominal solution that minimizes the worst-case expected violation. We note that this solution's worst-case expected violation is approximately half that of the initial nominal solution found. In practice, the proposed method can thus also be used to simply select the most robust nominal solution out of a set of possible optimal nominal solutions.

For left-hand side uncertainty, the standard robust counterparts of five of the considered problems are infeasible. Similar to how we treat such problems in Section 5.5, we include the IECV for these problems in Table 4. Figure 5 shows the characteristics of the solutions we found for **bnl1** and **perold**. For **bnl1** in particular, decreasing the first half of constraint violation hardly increases the optimal objective value, while there is a steep increase in optimal objective value close to the lowest possible worst-case expected violation.

When we do assume pair-wise independence for ζ_{ij} , the exact value of the worst-case expected violation can be computed with the methodology outlined in Section 3. The main reason to use a bound instead stems from the computational intractability of this method, as this requires $m \cdot 3^n$ evaluations in general. Therefore, an exact analysis for the ambiguity set with independence is only possible for very small problems. We can, however, estimate the worst-case expected violation for the pair-wise independence ambiguity set as we know the worst-case distribution. Sampling from this distribution is easy, as we assume all ζ_{ij} to be independent and Section 3 tells us the worst-case distribution is a discrete distribution on only three possible values. In general, we sample 10,000 observations (violations) and estimate the worst-case expected violation by their average.

We estimate the worst-case expected constraint violation for this ambiguity set for

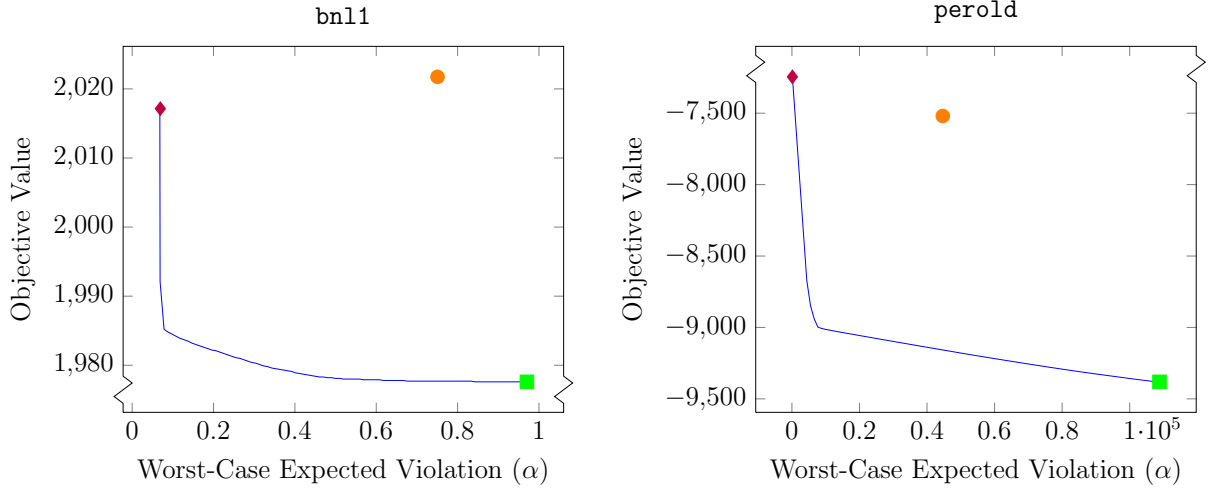


Figure 5: The solutions found by our approach with left-hand side uncertainty for **bn11** and **perold**, shown in blue. The green square indicates the nominal solution while the orange circle represents the solution found by shrinking the uncertainty set until we obtain feasibility for the standard robust counterpart. The purple diamond shows the solution found by minimizing the sum of slacks.

the nominal solution and the solutions corresponding to the IOV in Table 4. Table 1 illustrates the results for a representative set of problems. For these problems we observe that the worst-case expected violation for the ambiguity set with pair-wise independence is only occasionally higher, and if so never by much, than without assuming independence. This is true for every NETLIB problem, indicating that even if one is truly interested in an ambiguity set including pair-wise independence, omitting this assumption does little harm in terms of solution quality, while it does yield a computationally tractable problem.

Allowed violation for (6)	10%	25%	50%
blend	9.7	25.7	52.6
brandy	1.9	23.6	29.4
e226	7.6	21.3	50.0
greenbea	8.8	24.3	46.5
scfxm1	5.4	9.6	22.4

Table 1: Estimates of the actual allowed worst-case expected constraint violation in percentages when assuming pair-wise independence for the solutions that allow for 10%, 25% and 50% of the nominal solution’s worst-case expected violation without this assumption.

5.4 Removing Nominal Feasibility

For some problems, violation is truly interchangeable between constraints. For a facility location problem, for example, where constraints require demand in different locations to be met, a decision maker might not care about which demand is unmet. When we drop the requirement of solutions being nominal feasible, we find that we can significantly improve the nominal solution in objective value or worst-case expected violation. In this section, we assume right-hand side uncertainty identical to the uncertainty assumed in Section 5.2. Figure 6 shows the characteristics of the solutions found by our approach as well as the solutions to (P4) with the nominal feasibility constraint omitted. First of all, it is clear that the equivalence of (P4) and (P8) truly depends on the assumption of nominal feasibility, as the solutions to (P4) clearly perform worse on objective value and worst-case expected violation. The two different problems in Figure 6 show that the magnitude of differences between solutions to both problems is problem dependent.

Moreover, especially for `israel`, there exist solutions with a much lower objective value but with the same worst-case expected violation. Similarly, there exist solutions with the same objective value but a much lower worst-case expected violation than the nominal solution. Table 2 lists the potential improvement in both objective value and worst-case expected violation for the five problems for which they are the largest. All numbers are reported as differences to the nominal solution in percentages. Note that these potential improvements come at the price of the correspond-

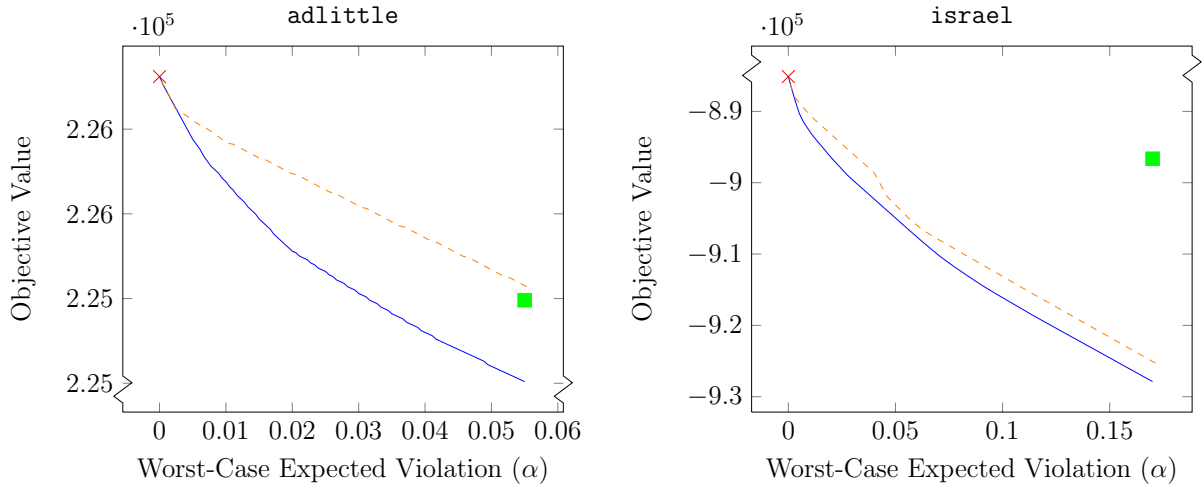


Figure 6: Solutions for `adlittle` and `israel` with right-hand side uncertainty, when nominal feasibility is not enforced. The solutions found by our approach are shown as the blue solid line and the solutions to (P4) as the orange dashed line. The red cross indicates the standard robust solution and the green square indicates the nominal solution.

Problem	Improvement	Improvement
	Objective value	Worst-case expected violation
agg	17.4	91.9
agg2	17.4	91.9
agg3	33.2	88.9
perold	4.8	91.9
pilot	8.6	96.0

Table 2: Potential improvement compared to the nominal solution in percentage when nominal feasibility is not enforced and right-hand side uncertainty is considered, $\epsilon = 0.01$ and $\delta = 0.5$.

ing solutions no longer being nominally feasible. The potential improvements are quite significant for these problems. Apparently, if one is willing to accept violation for the nominal scenario in some constraints, the total worst-case expected violation can be decreased by up to 90% without any cost or alternatively the objective value could potentially be improved 10% or more.

5.5 Worst-Case Violation

In considering the worst-case violation we consider a budget uncertainty set given by

$$U = \left\{ b \in \mathbb{R}^m : \left| \frac{b_i - \bar{b}_i}{\bar{b}_i} \right| \leq \epsilon, \quad i = 1, \dots, m, \quad \sum_{i=1}^m \left| \frac{b_i - \bar{b}_i}{\bar{b}_i} \right| \leq \Gamma m \epsilon \right\},$$

where we consider $\epsilon = 0.01$ and $\Gamma = 0.5$. Out of the 61 problems we considered, the reformulation suggested by Ardestani-Jaafari and Delage (2016) could not be solved within 30 minutes for 14 problems. It is important to note that many of the problems in the NETLIB library have numerous non-binding constraints included. For a high enough budget parameter Γ , this means that even though a budget constraint is present, all uncertain parameters in binding constraints can deviate from their nominal value maximally. Therefore, the uncertainty set is effectively a box uncertainty set on the binding constraints and thus (P4) is an equivalent problem that can be solved efficiently. In practice, however, the budget parameter should be chosen such that the budget constraint limits the effect that considering the worst-case scenario has on the solution, i.e., low enough that not every uncertain parameter in binding constraints can deviate from their nominal value.

6 Conclusion

In this paper we consider a new approach to find robust solutions to linear optimization problems with uncertain parameters. In particular, we attempt to alleviate the conservative nature of traditional robust optimization, which results in solutions that can be much worse in terms of objective value than the nominal solution. We find a wide array of solutions by allowing constraint violation up to a predefined maximum. The solutions found by our approach, together with the nominal and standard robust solution result in a good understanding of the trade-off between the objective value and the constraint violation for a problem.

In the resulting approach, we consider an ambiguity set based on the support, mean and mean-absolute deviation of uncertain parameters. This allows us to calculate the worst-case expected violation and bound it from above and leads to a tractable problem for both right- and left-hand side uncertainty. The approach is applied to all relevant NETLIB problems for both left- and right-hand side uncertainty. For many of these problems the array of solutions found gives great insight into the trade-off between objective value and worst-case expected violation. A selection of these solutions can also be found for problems for which the standard robust counterpart is infeasible. This allows decision makers to safeguard against at least part of the uncertainty for such problems and make better informed decisions with regard to this trade-off.

When one does not require the solution to be feasible for the nominal scenario, the potential improvements compared to the nominal solution can be enormous. For one problem, for example, a solution was found with the same objective value as the nominal solution of which the worst-case expected violation was only 5% of the nominal solution's worst-case expected violation.

Acknowledgments

The research of the first author was partly funded by the Netherlands Organisation for Scientific Research (NWO) Research Talent [Grant 406.17.511].

References

- Alvarez-Miranda, E., Ljubic, I., Raghavan, S., and Toth, P. (2015). The Recoverable Robust Two-Level Network Design Problem. *INFORMS Journal on Computing*, 27(1):1–19.
- Ardestani-Jaafari, A. and Delage, E. (2016). Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research*, 64(2):474–494.
- Baron, O., Milner, J., and Naseraldin, H. (2011). Facility location: A robust optimization approach. *Production and Operations Management*, 20(5):772–785.
- Ben-Tal, A., Bertsimas, D., and Brown, D. B. (2010). A soft robust model for optimization under ambiguity. *Operations Research*, 58(4-part-2):1220–1234.
- Ben-Tal, A., Boyd, S., and Nemirovski, A. (2006). Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Mathematical Programming*, 107(1-2):63–89.
- Ben-Tal, A., Brekelmans, R., Hertog, D. D., and Vial, J.-P. (2017). Globalized Robust Optimization for Nonlinear Uncertain Inequalities. *INFORMS Journal on Computing*, 29(2):350–366.
- Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust optimization*. Princeton University Press.
- Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805.
- Ben-Tal, A. and Nemirovski, A. (2000). Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88(3):411–424.
- Bertsimas, D., Nohadani, O., and Teo, K. M. (2010). Nonconvex Robust Optimization for Problems with Constraints. *INFORMS Journal on Computing*, 22(1):44–58.
- Bertsimas, D. and Sim, M. (2004). The price of robustness. *Operations Research*, 52(1):35–53.
- Chen, W., Sim, M., Sun, J., and Teo, C.-P. (2010). From cvar to uncertainty set: Implications in joint chance-constrained optimization. *Operations Research*, 58(2):470–485.
- El Ghaoui, L. and Lebret, H. (1997). Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications*, 18(4):1035–1064.
- Fiacco, A. V. and Kyparisis, J. (1986). Convexity and concavity properties of the optimal value function in parametric nonlinear programming. *Journal of Optimization Theory and Applications*, 48(1):95–126.
- Fischetti, M. and Monaci, M. (2009). Light robustness. In *Robust and online large-scale optimization*, pages 61–84. Springer.

- Gabrel, V., Lacroix, M., Murat, C., and Remli, N. (2014). Robust location transportation problems under uncertain demands. *Discrete Applied Mathematics*, 164:100–111.
- Goerigk, M. and Schöbel, A. (2016). Algorithm engineering in robust optimization. In *Algorithm engineering*, pages 245–279. Springer.
- Gorissen, B. L. and Den Hertog, D. (2013). Robust counterparts of inequalities containing sums of maxima of linear functions. *European Journal of Operational Research*, 227(1):30–43.
- Iancu, D. A. and Trichakis, N. (2014). Pareto Efficiency in Robust Optimization. *Management Science*, 60(1):130–147.
- Madansky, A. (1959). Bounds on the expectation of a convex function of a multivariate random variable. *The Annals of Mathematical Statistics*, 30(3):743–746.
- Mudchanatongsuk, S., Ordóñez, F., and Liu, J. (2008). Robust solutions for network design under transportation cost and demand uncertainty. *Journal of the Operational Research Society*, 59(5):652–662.
- Mutapcic, A. and Boyd, S. (2009). Cutting-set methods for robust convex optimization with pessimizing oracles. *Optimization Methods & Software*, 24(3):381–406.
- Pessoa, A. A. and Poss, M. (2015). Robust network design with uncertain outsourcing cost. *INFORMS Journal on Computing*, 27(3):507–524.
- Postek, K., Ben-Tal, A., Den Hertog, D., and Melenberg, B. (2018). Robust optimization with ambiguous stochastic constraints under mean and dispersion information. *Operations Research*, 66(3):814–833.
- Rockafellar, R. T. (1997). *Convex analysis*. Princeton University Press, Princeton, NJ.
- Schöbel, A. (2014). Generalized light robustness and the trade-off between robustness and nominal quality. *Mathematical Methods of Operations Research*, 80(2):161–191.
- Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2014). *Lectures on stochastic programming: modeling and theory*, volume 16. SIAM.
- Shapiro, A. and Nemirovski, A. (2005). On complexity of stochastic programming problems. In *Continuous optimization*, pages 111–146. Springer.
- Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157.
- Wiesemann, W., Kuhn, D., and Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376.

A Numerical Results for NETLIB Problems

	IOV				IECV		
	10.0%	25.0%	50.0%	100.0%	0.1%	0.5%	1.0%
25fv47	1.1	1.7	2.1	2.2	52.9	74.7	83.2
80bau3b	0.6	0.9	1.1	1.2	59.6	88.4	98.8
adlittle	0.2	0.3	0.5	0.6	58.5	95.1	100.0
afiro	0.2	0.5	1.0	1.0	52.2	73.4	100.0
agg	2.9	3.7	3.9	3.8	55.4	58.6	82.6
agg2	2.9	3.7	3.9	3.8	55.4	58.6	82.6
agg3	5.6	7.4	8.2	9.1	20.4	44.0	50.3
beaconfd	0.0	0.0	0.0	0.0	100.0	100.0	100.0
blend	0.2	0.5	0.8	1.0	24.3	74.9	100.0
bnl2	0.5	0.6	0.6	0.6	85.5	92.3	99.0
boeing					41.5	68.1	80.0
boeing2					32.0	71.3	72.5
brandy	0.0	0.0	0.1	0.1	100.0	100.0	100.0
capri	1.0	1.2	1.3	1.4	44.5	95.1	97.8
czprob	0.2	0.4	0.6	0.7	29.5	47.4	59.7
d2q06c	0.1	0.2	0.2	0.2	92.1	100.0	100.0
degen2					23.5	23.5	23.5
degen3					32.3	32.3	32.3
e226	0.9	1.2	1.4	1.5	50.9	86.4	97.1
etamacro	0.0	0.1	0.1	0.1	100.0	100.0	100.0
ffff800					74.6	84.2	86.0
finnis	1.5	2.7	3.3	3.4	56.9	67.9	76.5
ganges	0.1	0.3	0.4	0.5	40.1	96.4	100.0
gfrd-pnc	0.0	0.0	0.0	0.1	100.0	100.0	100.0
israel	1.1	1.3	1.3	1.3	85.0	91.6	91.3
lotfi	0.1	0.2	0.4	0.5	56.5	100.0	100.0
maros	0.2	0.3	0.4	0.4	79.6	80.1	90.9
nesm	0.0	0.0	0.0	0.0	100.0	100.0	100.0
perold	1.0	1.0	1.0	1.1	85.5	96.8	97.8
pilot	1.8	1.9	1.9	1.9	88.6	98.1	96.4
pilot4	0.1	0.1	0.1	0.1	91.3	100.0	100.0
pilot87	1.4	1.8	1.9	2.0	70.3	88.2	95.2
pilot_ja	0.7	0.7	0.8	0.8	95.4	99.2	100.0
pilot_we	0.3	0.5	0.7	0.7	60.0	78.4	92.0
sc105	0.4	0.7	0.9	1.0	46.4	85.5	100.0
sc205	0.7	0.9	1.0	1.0	75.1	92.7	97.9
sc50a	0.2	0.4	0.7	1.0	28.8	68.5	100.0
sc50b	0.2	0.4	0.7	1.0	19.0	65.3	100.0
scagr25	0.1	0.1	0.1	0.1	94.8	100.0	100.0
scagr7	0.0	0.0	0.1	0.1	64.1	100.0	100.0
scfxm1	0.1	0.2	0.2	0.2	87.2	98.2	100.0
scfxm2	0.1	0.2	0.2	0.2	62.7	95.2	100.0
scfxm3	0.1	0.2	0.2	0.2	60.8	94.1	100.0
scorpion	0.5	0.8	1.0	1.1	53.3	88.0	99.0
scrs8	0.5	0.9	1.0	1.0	70.5	90.8	99.9
sctap1	0.6	1.2	1.7	2.1	27.2	57.6	77.9
sctap2	0.4	0.7	0.9	1.2	30.3	76.4	96.8
sctap3	0.5	0.7	0.9	1.1	36.0	85.8	98.7
seba	0.1	0.2	0.3	0.4	61.5	100.0	100.0
share2b	0.5	0.8	1.0	1.0	57.7	89.9	99.7
ship04l	0.0	0.0	0.0	0.0	100.0	100.0	100.0
ship04s	0.0	0.0	0.0	0.0	100.0	100.0	100.0
ship08l	0.0	0.0	0.1	0.1	99.6	100.0	100.0
ship08s	0.0	0.0	0.1	0.1	93.3	100.0	100.0
ship12l	0.1	0.2	0.3	0.3	62.2	100.0	100.0
ship12s	0.1	0.2	0.3	0.4	61.2	100.0	100.0
sierra	0.0	0.0	0.0	0.0	83.6	100.0	100.0
stair	0.1	0.1	0.2	0.2	81.1	100.0	100.0
vtb_base	0.1	0.2	0.4	0.7	17.4	73.8	100.0
woodlp	0.3	0.7	1.3	2.7	3.7	18.3	36.6
woodw	0.2	0.6	1.0	1.3	33.9	68.0	88.9

Table 3: Characteristics of solutions found for right-hand side uncertainty, $\epsilon = 0.01$ and $\delta = 0.5$. All numbers are percentages relative to the nominal solution.

	IOV				IECV		
	10.0%	25.0%	50.0%	100.0%	0.1%	0.5%	1.0%
25fv47	0.3	0.3	0.4	0.4	93.7	100.0	100.0
80bau3b	0.8	1.5	2.1	2.3	23.0	69.0	79.7
adlittle	0.5	0.9	1.4	1.6	46.2	63.9	85.7
afiro	0.0	0.0	0.1	0.1	85.8	100.0	100.0
agg	0.1	0.1	0.1	0.1	98.6	99.8	100.0
agg2	0.1	0.1	0.1	0.1	98.6	99.8	100.0
agg3	0.2	0.2	0.2	0.2	39.3	100.0	100.0
blend	0.0	0.0	0.1	0.1	89.3	100.0	100.0
bnl1					64.4	92.7	92.9
bnl2	0.3	0.3	0.3	0.3	94.9	98.6	99.9
boeing2	0.0	0.0	0.0	0.1	100.0	100.0	100.0
brandy	0.0	0.0	0.0	0.0	100.0	100.0	100.0
capri	4.0	4.1	4.2	4.2	86.9	92.6	95.4
cycle	0.0	0.0	0.0	0.0	100.0	100.0	100.0
czprob	0.2	0.4	0.5	0.7	32.5	85.1	100.0
d2q06c					61.9	88.4	93.8
e226	2.7	2.8	2.8	2.9	88.9	91.7	93.9
etamacro	0.0	0.0	0.0	0.0	100.0	100.0	100.0
ffffff800	0.2	0.5	0.5	0.5	77.7	98.9	100.0
finnis	1.1	2.0	2.9	3.4	34.7	51.6	66.3
greenbea	4.9	5.1	5.1	5.1	83.6	94.0	95.0
greenbeb	0.1	0.2	0.2	0.2	97.9	100.0	100.0
israel	0.4	0.4	0.4	0.5	69.8	99.3	100.0
kb2	0.2	0.2	0.2	0.2	98.6	100.0	100.0
maros	0.5	1.0	1.5	1.8	29.1	60.9	81.1
nesm	0.1	0.4	0.7	0.9	40.9	73.5	100.0
perold					3.9	14.9	27.8
pilot					30.5	54.6	69.9
pilot4	1.1	1.8	2.7	3.6	21.2	37.5	53.1
pilot87					81.8	90.1	94.4
pilot_ja	1.3	1.8	2.7	3.6	13.6	33.6	51.5
pilot_we	1.4	2.0	2.5	2.7	29.6	67.6	84.2
pilotnov	2.0	2.0	2.0	2.0	99.9	99.9	99.9
scfxm1	0.1	0.2	0.2	0.3	47.0	100.0	100.0
scfxm2	0.1	0.2	0.2	0.3	51.2	100.0	100.0
scfxm3	0.1	0.2	0.2	0.3	51.8	100.0	100.0
scorpion	0.0	0.0	0.0	0.0	100.0	100.0	100.0
scrs8	0.0	0.0	0.0	0.0	100.0	100.0	100.0
ship04l	0.0	0.0	0.0	0.0	100.0	100.0	100.0
ship04s	0.0	0.0	0.0	0.0	100.0	100.0	100.0
ship08l	0.0	0.0	0.1	0.1	100.0	100.0	100.0
ship08s	0.0	0.0	0.1	0.1	91.4	100.0	100.0
ship12l	0.1	0.1	0.2	0.3	58.9	100.0	100.0
ship12s	0.1	0.2	0.2	0.3	58.9	100.0	100.0
stair	0.5	0.9	1.2	1.6	20.1	59.3	86.1
vtb_base	0.0	0.0	0.0	0.0	100.0	100.0	100.0

Table 4: Characteristics of solutions for left-hand side uncertainty, $\epsilon = 0.01$ and $\delta = 0.5$. All numbers are percentages relative to the nominal solution.