2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

A Scenario Approach for Estimating the Suboptimality of Linear Decision Rules in Two-Stage Robust Optimization

Michael J. Hadjiyiannis* and Paul J. Goulart† and Daniel Kuhn*

Abstract— **Robust dynamic optimization problems involving adaptive decisions are computationally intractable in general. Tractable upper bounding approximations can be obtained by requiring the adaptive decisions to be representable as linear decision rules (LDRs). In this paper we investigate families of tractable lower bounding approximations, which serve to estimate the degree of suboptimality of the best LDR. These approximations are obtained either by solving a dual version of the robust optimization problem in LDRs or by utilizing an inclusion-wise discrete approximation of the problem's uncertainty set. The quality of the resulting lower bounds depends on the distribution assigned to the uncertain parameters or the choice of the discretization points within the uncertainty set, respectively. We prove that identifying the best possible lower bounds is generally intractable in both cases and propose an efficient procedure to construct suboptimal lower bounds. The resulting instance-wise bounds outperform known worst-case bounds in the majority of our test cases.**

I. Introduction

Robust optimization is a powerful modeling paradigm for decision-making under uncertainty [1]. It is tailored to decision problems in which the distribution of the uncertain parameters is unknown except for its support. By definition, the support represents the range of all possible parameter realizations and is commonly referred to as the *uncertainty set*. Robust optimization models are designed to find the best decision in view of the worst-case realization of the uncertain parameters within their uncertainty set.

Classical *static* robust optimization models involve only *design decisions*, which are of here-and-now type and must be selected before any of the uncertain parameters are observed. Recently, *dynamic* robust optimization problems have attracted considerable interest. Such problems incorporate additional *adaptive decisions*, which are of wait-and-see type and can be selected after the uncertain parameters have been revealed. Thus, adaptive decisions are modeled as *decision rules*, that is, functions of the uncertain parameters.

Robust optimization problems involving adaptive decisions are generally computationally intractable [2], [3]. To reduce their complexity, methods for restricting the space of adaptive decision rules have been suggested, e.g. by restricting the decision rules to those with linear [2], piecewise linear [4], [5] or polynomial [6] structure. Because of their desirable scalability properties, LDRs enjoy the widest popularity. Indeed, the best LDR for a given robust optimization problem can typically be computed by solving a tractable linear or second-order cone program.

Unfortunately, LDRs may be severely suboptimal or even infeasible in the original optimization problem. In order to assess the appropriateness of LDRs, one thus needs to estimate their loss of optimality. We distinguish two complementary approaches for this purpose: *a priori* methods evaluate the worst-case approximation ratio of LDRs over a whole class of problems, while *a posteriori* methods estimate the approximation error for each problem instance individually.

We first review the existing a priori methods. Linear decision rules have been shown to be optimal for two-stage minmax problems with simplicial uncertainty sets and for certain one-dimensional robust control problems [7]. However, instances satisfying these idealized conditions are rarely encountered in practice. Under restrictive nonnegativity conditions on the problem data, it has also been shown that the worst-case approximation ratio of LDRs in two-stage minmax problems with m linear constraints is of the order $\Omega(\sqrt{m})$ [3]. Even though it is occasionally tight, this worst-case performance bound is too pessimistic for the vast majority of problem instances.

Thus, there is considerable merit in developing good a posteriori methods. Instance-wise upper and lower bounds on multi-stage *stochastic* programs can be obtained by solving both the original problem and its dual in LDRs; see [8]. The gap between the bounds provides an a posteriori measure for the suboptimality of LDRs. This approach is directly applicable to robust optimization problems if a probability distribution is assigned to the uncertain parameters. However, in this case the dual (lower) bound becomes non-unique as it depends on that distribution (which can be chosen freely from amongst distributions with appropriate support).

In this paper we demonstrate that the quality of the dual LDR bound for robust problems is highly sensitive to the choice of the distribution governing the uncertain parameters. We show that the best (maximum) lower bound can be achieved by a *discrete* distribution. However, we further show that finding this distribution is as hard as solving the original (intractable) problem. A different class of tractable lower bounds is obtained by replacing the original uncertainty set with a finite scenario set of a prescribed cardinality. Again, finding the best scenario set is generically hard. We therefore propose an efficient method in which scenarios are constructed from the Lagrange multipliers associated with the primal LDR problem. Next, we establish a link between the dual LDR method and the scenario approach.

* Department of Computing, Imperial College London
† Department of Aeronautics, Imperial College London
{mjh205,p.goulart,dkuhn}@imperial.ac.uk
This work was supported by EPSRC Grant EP/H0204554/1

This analysis allows us to identify new problem classes for which LDRs are optimal. Extensive numerical experiments demonstrate that our scenario-based lower bound consistently outperforms the known a priori bounds as well as the dual LDR bound associated with naive distributional choices.

This paper is structured as follows. Section II introduces the problem formulation and reviews an approximate solution method based on LDRs. Sections III and IV describe bounding techniques for estimating the suboptimality of LDRs by using dual LDRs and a discrete approximation of the uncertainty set, respectively. The performance of the resulting bounds is analyzed in Section V on a set of randomly generated test problems.

Notation: The optimal value of any optimization problem $\mathscr{P}$ is denoted by $\mathscr{P}^*$. For matrices $S, T \in \mathbb{R}^{(k+1)\times l}$ and a proper cone $\mathscr{K} \subseteq \mathbb{R}^{k+1}$, the relation $S \succeq_{\mathscr{K}} T$ ($S \preceq_{\mathscr{K}} T$) indicates that the columns of $S-T$ ($T-S$) are included in the cone $\mathscr{K}$. The dual cone of $\mathscr{K}$ is denoted by $\mathscr{K}^*$. Moreover, the kernel of a matrix A is denoted $\ker(A)$. Finally, we let $A^{(i)}$ be the i-th row and $A_{(i)}$ the i-th column of A.

II. Problem Statement

We study linear two-stage robust optimization problems with affine right hand side uncertainty and a minmax objective. Such problems can be represented as

$$\begin{array}{ll} \inf & c^\top x + \sup\limits_{\xi\in\Xi} d^\top y(\xi) \\ \text{s.t.} & Ax + By(\xi) \le C\xi \quad \forall \xi \in \Xi, \end{array} \tag{$\mathscr{P}$}$$

where $\Xi \subseteq \mathbb{R}^{k+1}$ represents the uncertainty set. The first-stage decision $x \in \mathbb{R}^{n_1}$ is a rigid design decision, and the second-stage decision y is a fully adaptive decision rule, that is, a continuous function from $\mathbb{R}^{k+1}$ to $\mathbb{R}^{n_2}$. We denote by m the number of constraints. Without much loss of generality, we assume that $\mathscr{P}^*$ is finite and that the minimum is attained by an optimal decision (x^*, y^*). Moreover, we assume that the uncertainty set Ξ is defined as $\Xi := \left\{\xi \in \mathscr{K} \ : \ e_0^\top \xi = 1\right\}$, where $\mathscr{K} \subseteq \mathbb{R}^{k+1}$ is a proper cone and e_0 is the first canonical basis vector in $\mathbb{R}^{k+1}$. We also assume that Ξ is non-empty and bounded.

Remark 1: Introducing a degenerate uncertain variable ξ_0 that is equal to 1 on Ξ allows us to express any affine function of the non-degenerate uncertain parameters $(\xi_1, \ldots, \xi_m)$ on Ξ in a compact way as a linear function of $\xi = (\xi_0, \ldots, \xi_m)$.

Problem $\mathscr{P}$ is computationally intractable [2, Theorem 2.2], involving an infinite number of constraints and decision variables. Therefore, finding a suboptimal solution necessitates a trade-off between accuracy and tractability, usually in the form of a restriction on the structure of the adaptive decision rules in $\mathscr{P}$. In this paper, we investigate such an approximation where the second-stage adaptive decision rule y is restricted to be a linear function of ξ, i.e., $y(\xi) = Y\xi$ for some matrix $Y \in \mathbb{R}^{n_2\times(k+1)}$. The problem of identifying an optimal LDR can be represented as

$$\begin{array}{ll} \inf & c^\top x + \sup\limits_{\xi\in\Xi} d^\top Y\xi \\ \text{s.t.} & Ax + BY\xi \le C\xi \quad \forall \xi \in \Xi, \end{array} \tag{$\mathscr{U}$}$$

where the second-stage decision is now encoded by the matrix Y instead of the continuous function y. Problem $\mathscr{U}$ constitutes a linear robust optimization problem with semi-infinite constraints. By using robust optimization techniques [2], it can be reformulated as a conic optimization problem of the form

$$\begin{array}{ll} \inf & c^\top x + t \\ \text{s.t.} & \left(d^\top Y - te_0^\top\right)^\top \preceq_{\mathscr{K}^*} 0 \\ & \left(Axe_0^\top + BY - C\right)^\top \preceq_{\mathscr{K}^*} 0. \end{array} \tag{$\tilde{\mathscr{U}}$}$$

Problem $\mathscr{U}$ and its equivalent conic reformulation $\tilde{\mathscr{U}}$ are derived via a restriction of the set of admissible second-stage decisions in $\mathscr{P}$. In general, an optimal fully adaptive decision rule y^* is continuous piecewise linear in ξ, with a possibly exponential number of pieces [9]. Restricting y to be linear in ξ introduces an optimality gap between $\mathscr{P}$ and $\mathscr{U}$ with $\mathscr{P}^* \le \mathscr{U}^* = \tilde{\mathscr{U}}^*$. The loss of performance due to the use of LDRs varies greatly according to the problem data. Identifying a bound on this optimality gap requires the derivation of a lower bound for $\mathscr{P}^*$ tailored to the specific problem instance. In the remainder of this paper we demonstrate how one can derive such problem-specific lower bounds in an efficient manner.

Dual variables and the robust counterpart $\tilde{\mathscr{U}}$

Before we discuss the first lower bound, we comment on the relationship between the semi-infinite problem $\mathscr{U}$ and its conic equivalent $\tilde{\mathscr{U}}$ and demonstrate how one can derive $\tilde{\mathscr{U}}$ from $\mathscr{U}$ using two distinct approaches. The first approach relies on the following lemma, which captures the essence of robust optimization:

Lemma 2.1 ([10]): For any $\sigma \in \mathbb{R}^{k+1}$ we have

$$\sigma^\top \xi \ge 0 \quad \forall \xi \in \Xi \iff \sigma \in \mathscr{K}^*,$$

where $\mathscr{K}$ is the cone generated by Ξ, and $\mathscr{K}^*$ its dual cone. Problem $\tilde{\mathscr{U}}$ can be derived from $\mathscr{U}$ via application of Lemma 2.1, which has the effect of replacing the semi-infinite linear inequality constraints in $\mathscr{U}$ with a set of finite dimensional $\mathscr{K}^*$-conic constraints.

We suggest now an alternative derivation based on Lagrangian functions, which provides a different insight into the relationship between $\mathscr{U}$ and $\tilde{\mathscr{U}}$. Let $\mathscr{M}_+^d$ be the space of all d-dimensional nonnegative Borel measures on Ξ. Let f be any continuous function from $\mathbb{R}^{k+1}$ to $\mathbb{R}^d$. Then we have

$$\sup_{\psi\in\mathscr{M}_+^d} \int_\Xi f(\xi)^\top \psi(\mathrm{d}\xi) = \left\{ \begin{array}{ll} \infty & \text{if } \exists \xi \in \Xi : f(\xi) > 0, \\ 0 & \text{otherwise.} \end{array} \right. \tag{1}$$

Let $q \in \mathscr{M}_+^m$ and $z \in \mathscr{M}_+$ be m-dimensional and one-dimensional non-negative Borel measures on Ξ, respectively. By using (1), we can derive the Lagrangian of $\mathscr{U}$ using q and z as the dual variables for the semi-infinite inequality constraints, i.e.,

$$\begin{aligned} L_{\mathscr{U}}(x, Y, t; q, z) = \quad & c^\top x + t + \left(d^\top Y - te_0^\top\right) \textstyle\int_\Xi \xi z(\mathrm{d}\xi) \\ & + \mathrm{tr}\left[\left(Axe_0^\top + BY - C\right) \textstyle\int_\Xi \xi q^\top(\mathrm{d}\xi)\right]. \end{aligned}$$

The emergence of Borel measures as dual variables in the Lagrangian is a consequence of the semi-infinite nature of the constraints in $\mathscr{P}$. However, these dual Borel measures only affect the Lagrangian through their first moments. Consider therefore the variable substitution $\Lambda = \int_\Xi \xi q^\top(\mathrm{d}\xi)$ and $\lambda = \int_\Xi \xi z(\mathrm{d}\xi)$, where the new variables Λ and λ are the first moments of the Borel measures q and z, respectively. The non-negativity constraints $q \in \mathscr{M}_+^m$ and $z \in \mathscr{M}_+$ translate to the new variables as $(\lambda, \Lambda) \succeq_{\mathscr{K}} 0$. Substituting these new variables into $L_{\mathscr{U}}$ yields:

$$\begin{aligned} L_{\mathscr{U}} &= c^\top x + t + \left(d^\top Y - t e_0^\top\right)\lambda + \operatorname{tr}\left[(A x e_0^\top + BY - C)\Lambda\right] \\ &= L_{\tilde{\mathscr{U}}}(x, Y, t; \lambda, \Lambda), \end{aligned} \tag{2}$$

which is precisely the Lagrangian of $\tilde{\mathscr{U}}$, where Λ and λ are the dual variables of the conic inequality constraints.

The preceding line of argument provides an alternate perspective on how the conic problem $\tilde{\mathscr{U}}$ relates to the semi-infinite problem $\mathscr{U}$. The established interpretation dictates that problem $\tilde{\mathscr{U}}$ is derived from its semi-infinite equivalent via a manipulation of the constraints of $\mathscr{U}$ that utilizes convex duality arguments, through the mechanism of Lemma 2.1. This alternative derivation can be interpreted as a simple variable substitution in the Lagrangians, where the dual Borel measures in $L_{\mathscr{U}}$ are substituted by their respective moments to obtain the Lagrangian $L_{\tilde{\mathscr{U}}}$.

III. Dual LDR Bound

The problem of deriving a lower bound for two-stage decision problems was addressed successfully in the stochastic programming setting [8], where a lower bound was proposed for linear problems similar to $\mathscr{P}$, but with expectation objective and a prescribed probability distribution for ξ. The bound derived in [8] employed a dualization of the original (infinite) problem, followed by a restriction of the dual variables to be linear in ξ. The resulting semi-infinite problem is a relaxation of $\mathscr{P}$ with expectation constraints depending on a probability distribution $\mathbb{P}$ [8]:

$$\begin{array}{ll} \inf & c^\top x + t \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}\left[\left(d^\top y(\xi) - t\right)\xi^\top\right] \leq 0 \\ & \mathbb{E}_{\mathbb{P}}\left[\left(Ax + By(\xi) - C\xi\right)\xi^\top\right] \leq 0. \end{array} \tag{$\mathscr{L}(\mathbb{P})$}$$

Even though $\mathscr{L}(\mathbb{P})$ involves functional decision variables and is therefore seemingly intractable, it has been shown that under a mild strict feasibility condition, $\mathscr{L}(\mathbb{P})$ can be reformulated as a tractable conic optimization problem [8]. Let $\mathscr{C}$ be the set of all probability distributions supported on Ξ. One can immediately verify that $\mathscr{L}(\mathbb{P})$ provides a lower bound on $\mathscr{P}$ for any possible distribution $\mathbb{P} \in \mathscr{C}$:

Proposition 3.1: For any probability distribution $\mathbb{P} \in \mathscr{C}$, we have $\mathscr{L}^*(\mathbb{P}) \leq \mathscr{P}^* \leq \mathscr{U}^*$.

Any solution that is feasible in $\mathscr{P}$ will of course also satisfy the less restrictive expectation constraints in $\mathscr{L}(\mathbb{P})$, regardless of the probability distribution used. However, the choice of a distribution $\mathbb{P}$ has a central role in the performance of the bound $\mathscr{L}^*(\mathbb{P})$.

Constraints in the original problem $\mathscr{P}$ corresponding to uncertainties with higher probability mass under $\mathbb{P}$ are more likely to be satisfied by the optimal solution of $\mathscr{L}(\mathbb{P})$. In the absence of any information about the distribution $\mathbb{P}$, as is the case with $\mathscr{P}$, we are free to select any distribution from $\mathscr{C}$ to be used in $\mathscr{L}(\mathbb{P})$.

In the extreme case, one can adopt a Dirac distribution from $\mathscr{C}$ that concentrates probability mass on a single uncertainty realization $\xi \in \Xi$. A solution to $\mathscr{L}(\mathbb{P})$ will then only need to satisfy the constraints corresponding to that single uncertainty realization, i.e., the problem becomes deterministic. In the event that such constraints are not binding, the optimal value of $\mathscr{L}(\mathbb{P})$ is $-\infty$, and the lower bound is trivial.

Consequently, if one wishes to obtain a useful lower bound from $\mathscr{L}(\mathbb{P})$, then the distribution $\mathbb{P}$ must be carefully chosen, ideally from the set of worst-case distributions

$$\mathscr{C}_w := \operatorname*{arg\,max}_{\mathbb{P} \in \mathscr{C}} \mathscr{L}(\mathbb{P}),$$

which can be shown to be non-empty. Selecting any distribution $\mathbb{P}_w \in \mathscr{C}_w$ ensures that $\mathscr{L}^*(\mathbb{P}_w)$ gives the best possible lower bound for $\mathscr{P}$ among all dual LDR bounds. Unfortunately, finding an element $\mathbb{P}_w \in \mathscr{C}_w$ and evaluating $\mathscr{L}^*(\mathbb{P}_w)$ is no less difficult than computing the optimal value of $\mathscr{P}$:

Theorem 3.2: The set $\mathscr{C}_w$ is non-empty, and for any worst-case distribution $\mathbb{P}_w \in \mathscr{C}_w$, the lower bounding problem $\mathscr{L}(\mathbb{P}_w)$ satisfies $\mathscr{L}^*(\mathbb{P}_w) = \mathscr{P}^*$. Thus, computing the best possible dual LDR bound is as hard as computing the optimal value of $\mathscr{P}$.

Proof: First define the optimal second-stage cost of $\mathscr{P}$, which depends parametrically on the first-stage decision x and the uncertainty realization ξ, i.e.,

$$\begin{array}{rll} Q(x,\xi) := & \min_y & d^\top y \\ & \text{s.t.} & Ax + By \leq C\xi. \end{array} \tag{3}$$

The optimal value function Q is known to be convex and continuous on its effective domain. Problem $\mathscr{P}$ can now be expressed as the following minmax problem:

$$\mathscr{P} = \inf\left[c^\top x + \sup_{\xi \in \Xi} Q(x,\xi)\right].$$

Let x^* be an optimal first-stage decision and Ξ_w be the set of worst-case scenarios for the corresponding optimal second-stage parametric cost,

$$\Xi_w := \operatorname*{arg\,max}_{\xi \in \Xi} Q(x^*, \xi). \tag{4}$$

Note that Ξ_w is non-empty since Q is continuous on its effective domain and Ξ is compact. For any scenario $\xi_w \in \Xi_w$ and any $x \in \mathbb{R}^{n_1}$, we have $Q(x, \xi_w) \geq Q(x^*, \xi_w)$. Now let δ_w be any Dirac distribution concentrating mass on a single element $\xi_w \in \Xi_w$. Then, we find

$$\begin{aligned} \mathscr{L}^*(\delta_w) &= \mathrm{inf_x}\ \mathrm{c^\top x + Q(x, \xi_w)} \\ &= c^\top x^* + Q(x^*, \xi_w) = \mathscr{P}^*. \end{aligned}$$

Thus, $\mathscr{C}_w$ is non-empty, and any $\mathbb{P}_w \in \mathscr{C}_w$ results in an exact lower bound $\mathscr{L}^*(\mathbb{P}_w) = \mathscr{P}^*$. Additionally, since the problem $\mathscr{L}(\delta_w)$ is actually finite and therefore a tractable convex cone program, finding a Dirac distribution $\delta_w \in \mathscr{C}_w$ has the same complexity as finding the optimal value of $\mathscr{P}$. ■

Theorem 3.2 has two important implications. First, finding the best dual LDR bound is actually as hard as deriving the optimal value of the original problem $\mathscr{P}$. Using any lower bound $\mathscr{L}^*(\mathbb{P})$ associated with a distribution $\mathbb{P} \notin \mathscr{C}_w$ will give only a conservative estimate of the suboptimality of LDRs.

Second, the proof of Theorem 3.2 illustrates that there exist specific uncertainty realizations, namely the elements of Ξ_w, from which Dirac distributions $\mathbb{P}_w$ can be constructed that provide tight lower bounds $\mathscr{L}^*(\mathbb{P}_w) = \mathscr{P}^*$.

As a result, a different method of bounding $\mathscr{P}$ can be designed, which is inspired by the existence of the set Ξ_w and relies on a discrete approximation of the uncertainty set.

IV. Scenario Based Bounds

We propose a procedure in which we identify a finite discrete subset $Z \subset \Xi$ and solve $\mathscr{P}(Z)$, a variant of the problem $\mathscr{P}$ in which the uncertainty set Ξ is replaced by Z. This results in a solution that is robust only with respect to a subset of all scenarios $\xi \in \Xi$, namely the elements of Z. Problem $\mathscr{P}(Z)$ is thus finite, and its solution provides a lower bound on $\mathscr{P}^*$.

Theorem 4.1: For any finite subset $Z \subset \Xi$, problem $\mathscr{P}(Z)$ provides a lower bound on $\mathscr{P}$ that can be obtained via the solution of a finite linear program. Moreover, for any distribution $\mathbb{P}_z \in \mathscr{C}$ supported on Z, $\mathscr{P}(Z)$ provides a tighter lower bound on $\mathscr{P}$ than $\mathscr{L}(\mathbb{P}_z)$, i.e., $\mathscr{L}^*(\mathbb{P}_z) \leq \mathscr{P}^*(Z) \leq \mathscr{P}^*$.

We omit the proof of this theorem for the sake of brevity. The quality of this lower bound depends on the choice of Z, in a manner equivalent to the dependence of $\mathscr{L}(\mathbb{P})$ on the choice of $\mathbb{P}$. The next result follows from the preceding discussion and is a generalization of Theorem 3.2.

Theorem 4.2: There exists a non-empty set $\Xi_w \subseteq \Xi$, so that for any $Z \subseteq \Xi$ with $Z \bigcap \Xi_w \neq \emptyset$, we have $\mathscr{P}^*(Z) = \mathscr{P}^*$.

We note that Theorem 4.2 is reminiscent of Theorem 3.2. While Theorem 3.2 relates to the existence of worst-case discrete distributions that maximize the dual LDR bound $\mathscr{L}^*(\mathbb{P})$, Theorem 4.2 guarantees the existence of finite worst-case scenario sets Z that maximize the discrete approximation bound $\mathscr{P}^*(Z)$. Of course, finding a worst-case scenario set, or equivalently an element of Ξ_w is at least as hard as computing $\mathscr{P}^*$ and thus is itself intractable. Thus, one must resort to other methods for choosing the set Z.

A naïve approach is to let Z be the set of all extreme points of Ξ. It is known that Ξ_w contains at least one of these extreme points [3], and therefore such Z is guaranteed to yield a tight bound.

Proposition 4.3: Let $Z_v := \operatorname{ext}(\Xi)$ be the set of extreme points of Ξ. Then $\Xi_w \bigcap Z_v \neq \emptyset$ and $\mathscr{P}^*(Z_v) = \mathscr{P}^*$.

If Ξ is strictly convex, then Z_v has infinite cardinality, and $\mathscr{P}(Z_v)$ remains an infinite problem. In the particular case where Z_v is finite, e.g. for polyhedral uncertainty sets, it is in principle possible to enumerate the vertices of Ξ. However, $\mathscr{P}(Z_v)$ is still generally intractable.

Remark 2: It is also possible to generate the discrete set Z by drawing finitely many samples from any distribution $\mathbb{P} \in \mathscr{C}$ supported on Ξ. The resulting scenario problem $\mathscr{P}(Z)$ yields a stochastic lower bound for $\mathscr{P}$ whose quality depends on the number of samples and the choice of $\mathbb{P}$. Scenario problems of this type have been investigated in [11], [12].

The Critical Set Δ

Theorem 4.2 guarantees that pruning the uncertainty set from Ξ to any non-empty subset of Ξ_w does not alter the optimal value of $\mathscr{P}$, even though such a reduction amounts to a relaxation of the constraints in $\mathscr{P}$. However, finding such a subset is hard.

We therefore turn to the binding uncertainty realizations in the LDR problem $\mathscr{U}$ as a proxy, i.e., we propose to use a finite scenario set Δ, which is derived from $\mathscr{U}$ in a similar way as Ξ_w is derived from $\mathscr{P}$. Such a set can be constructed efficiently and enjoys properties with respect to the LDR problem $\mathscr{U}$ similar to those of Ξ_w with respect to the original problem $\mathscr{P}$.

Theorem 4.4: There exists a non-empty set $\mathbb{D} \subseteq 2^{\Xi}$, where each $\Delta \in \mathbb{D}$ is a finite subset of Ξ with $|\Delta| \leq m+1$ and satisfies $\mathscr{U}^*(\Delta) = \mathscr{U}^*$. Here, $\mathscr{U}(Z)$ denotes a variant of problem $\mathscr{U}$ in which Ξ is replaced by Z. One particular set $\Delta \in \mathbb{D}$ can be constructed efficiently via the solution of the dual of the problem $\tilde{\mathscr{U}}$.

Proof: See Appendix. ■

We propose to use $\mathscr{P}^*(\Delta)$ for some $\Delta \in \mathbb{D}$ as a lower bound on $\mathscr{P}^*$. The motivation for this choice is predicated on the reasonable expectation that there may exist at least one worst-case uncertainty realization for $\mathscr{P}$ that is also a worst-case realization for $\mathscr{U}$, i.e., that there exists $\Delta \in \mathbb{D}$ such that $\Delta \bigcap \Xi_w \neq \emptyset$.

In cases where the above assumption holds and $\Delta \in \mathbb{D}$ is chosen correctly, $\mathscr{P}^*(\Delta)$ will be equal to the true optimal value of $\mathscr{P}$, and the optimality gap $\mathscr{U}^* - \mathscr{P}^*(\Delta)$ will be an exact characterization of the suboptimality of LDRs. On the other hand, our approach can return a conservative optimality gap for one of two reasons:

i) There does not exist any $\Delta \in \mathbb{D}$ with $\Delta \bigcap \Xi_w \neq \emptyset$. Such a situation is illustrated in Example V-A.
ii) There does exist some $\Delta \in \mathbb{D}$ with $\Delta \bigcap \Xi_w \neq \emptyset$, but $|\mathbb{D}| \geq 2$ and there is no mechanism of selecting the right one. In the numerical examples of V-B, we encountered several problem instances for which $\mathscr{P}^*(\Delta) < \mathscr{U}^*$ when in fact $\mathscr{P}^* = \mathscr{U}^*$.

Despite this limitation, we have found the approach to perform well in practice and to reliably identify problem instances for which LDRs are optimal.

Furthermore, properties of a given set $\Delta \in \mathbb{D}$ can in certain circumstances be used to identify immediately that LDRs are optimal without computing a lower bound.

Theorem 4.5 (Optimality of LDRs): If there exists a $\Delta \in \mathbb{D}$ whose elements are linearly independent, then LDRs are optimal for the associated instance of $\mathscr{P}$, and the optimality gap derived from such a Δ is zero, i.e.,

$$\mathscr{P}^*(\Delta) = \mathscr{P}^* = \mathscr{U}^*(\Delta) = \mathscr{U}^*.$$

Proof: The rightmost equality follows from Theorem 4.4. The remainder of the proof relies on the fact that LDRs are optimal for simplicial uncertainty sets [3]. If there exists a Δ with linearly independent elements, then the convex hull Δ_c of Δ constitutes a (maybe degenerate) simplex. As Δ contains all the extreme points of Δ_c, we have $\mathscr{U}^*(\Delta_c) = \mathscr{U}^*(\Delta)$ and $\mathscr{P}^*(\Delta_c) = \mathscr{P}^*(\Delta)$. Since LDRs are optimal for Δ_c, it follows that $\mathscr{U}^*(\Delta_c) = \mathscr{P}^*(\Delta_c)$. ■

Remark 3: Theorems 4.4 and 4.5 together provide a convenient method for establishing the optimality of LDRs for a specific instance of $\mathscr{P}$ without explicitly calculating a lower bound. One only needs to construct Δ via the solution of the dual of $\tilde{\mathscr{U}}$ and check whether its elements are linearly independent.

Corollary 4.6: Let $\mathbb{P}_\Delta$ be any probability distribution supported on Δ. Under the assumptions of Theorem 4.5, we find

$$\mathscr{L}^*(\mathbb{P}_\Delta) = \mathscr{P}^*(\Delta) = \mathscr{P}^* = \mathscr{U}^*(\Delta).$$

Thus, the lower bound $\mathscr{L}^*(\mathbb{P}_\Delta)$ also certifies the optimality of LDRs in these situations.

Efficient Identification of Worst-Case Scenarios

Despite the general intractability result of Theorem 3.2, it is sometimes possible to compute an element of Ξ_w.

Theorem 4.7: Suppose that $\ker(B^\top) \subseteq \ker(C^\top)$. Then,

$$\Xi_w = \arg\max_{\xi \in \Xi} -\mu^\top C \xi, \tag{5}$$

where μ is any vector satisfying $\mu \geq 0$ and $B^\top \mu = -d$.

Proof: Consider the optimal second-stage cost $Q(x,\xi)$ defined in (3). By linear programming duality, we obtain

$$Q(x,\xi) = \mu^\top(Ax - C\xi) + \begin{array}{ll} \max_p & p^\top(Ax - C\xi) \\ \text{s.t.} & B^\top p = 0 \\ & p \geq -\mu. \end{array} \tag{6}$$

Whenever the kernel condition holds, the maximization term in the above expression is independent of ξ. Recalling the definition (4) of Ξ_w we have

$$\Xi_w = \arg\max_{\xi \in \Xi} \; Q(x^*, \xi) = \arg\max_{\xi \in \Xi} \; \mu^\top C \xi,$$

and thus an element of Ξ_w can be found by solving a convex optimization problem. ■

Similar results can be obtained whereby alternative assumptions to those in Theorem 4.7 ensure that solutions to (5) are contained in Ξ_w, and consequently that $\mathscr{P}$ is efficiently solvable. For example one can replace the assumption that $\ker(B^\top) \subseteq \ker(C^\top)$ with the assumption that both $A = 0$ and $C^\top \ker(B^\top) \subseteq \mathscr{K}^*$; in this case the maximization term in (6) vanishes for all possible x and $\xi \in \Xi$, and thus Theorem 4.7 still holds. Such a situation occurs in Example V-A.

V. Numerical Results

Throughout this section, we compute lower bounds on $\mathscr{P}$ by solving $\mathscr{P}(Z)$ where: $Z = \Delta$ is the set defined in Theorem 4.4; $Z = \{\xi_m\}$ is a single scenario set where ξ_m is chosen via a solution of (5); and $Z = Z_v$ is the set of vertices of the support.[1] For completeness, we also provide a set of dual LDR bounds $\mathscr{L}^*(\mathbb{P})$, parametrized by three different uncertainty distributions: $\mathbb{P} = \mathbb{P}_\Xi$ is the uniform distribution on the support Ξ; $\mathbb{P} = \mathbb{P}_\Delta$ is the uniform distribution on the set Δ; and $\mathbb{P} = \mathbb{P}_{Z_v}$ is the uniform distribution on the set of vertices Z_v of the support. The estimated optimality gaps provided by each bound are then calculated as the difference between $\tilde{\mathscr{U}}^*$ and the different lower bounds described above, whilst the percentage gaps correspond to that difference divided by the respective lower bound.

[1]Recall from Proposition 4.3 that $\mathscr{P}^*(Z_v) = \mathscr{P}^*$, so that $\mathscr{U}^* - \mathscr{P}^*(Z_v)$ measures the true degree of suboptimality of LDRs.

TABLE I
BOUND COMPARISON FOR TEMPORAL NETWORK EXAMPLE

Problem	Opt. Value	Gap	% Gap
$\tilde{\mathscr{U}}$	2.00	-	-
$\mathscr{P}$	1.71	0.29	16%
$\mathscr{P}(\{\xi_m\})$	1.71	0.29	16%
$\mathscr{P}(\Delta)$	1.50	0.50	33%
$\mathscr{L}(\mathbb{P}_\Delta)$	1.40	0.60	43%
$\mathscr{L}(\mathbb{P}_\Xi)$	1.25	0.75	60%

A. Temporal Networks Example

We investigate a specific instance of a temporal network problem [13], for which LDRs are known to be suboptimal.

$$\begin{array}{ll} \inf & \sup_{\xi \in \Xi} y_2(\xi) \\ \text{s.t.} & \left.\begin{array}{l} y_1(\xi) \geq \max(\xi_1, 1-\xi_1) \\ -y_2(\xi) \geq y_1(\xi) + \max(\xi_2, 1-\xi_2) \end{array}\right\} \forall \xi \in \Xi \end{array} \tag{7}$$

There are two fully adaptive one-dimensional decisions y_1 and y_2 but no design decisions. The uncertainty set is $\Xi := \left\{\xi \in \mathbb{R}^3 \;:\; \xi_1 = 1, (\xi_1 - \frac{1}{2})^2 + (\xi_2 - \frac{1}{2})^2 \leq (\frac{1}{2})^2\right\}$. The optimal value of (7) can be calculated analytically [13]. Furthermore, one can find a worst-case scenario in Ξ_w by solving (5). The results for the various bounds are shown in Table I.

B. Randomly Generated Instances

We assess the performance of the different lower bounds based on a sample of randomly generated instances of $\mathscr{P}$ with fixed dimensions $k = 16$, $m = 16$, $n_1 = 3$ and $n_2 = 5$. The instances are generated to adhere to the assumptions described in [3], where it has been shown that $\mathscr{U}^*/\mathscr{P}^*$ is bounded below by $\Omega(\sqrt{m}) = 4$ in the worst case. This means that LDRs can incur a percentage gap of up to 300%. The problem instances are generated as follows:

- A constant symmetric uncertainty set Ξ is defined as $\Xi := \left\{(1,\xi)^\top \in \mathbb{R}^{k+1} \;:\; \|\xi\|_p \leq 1\right\}$ for $p \in \{1, 2, \infty\}$.
- The elements of the matrices A and B are uniformly distributed on the interval $[-5, 5]$.
- The matrix C is randomly generated so that $C^\top \succeq_{\mathscr{K}^*} 0$, which guarantees that $C\xi \geq 0$ for all $\xi \in \Xi$.
- A vector $\mu \in \mathbb{R}^m$ is randomly generated according to a uniform distribution on $[0,1]^m$. The vector μ is used to construct the cost vectors $c = -A^\top \mu$ and $d = -B^\top \mu$, ensuring that $c \geq 0$ and $d \geq 0$. The nonnegativity restrictions are necessitated by the assumptions in [3]. The existence of a μ satisfying the above is required by the postulated dual feasibility of the original problem $\mathscr{P}$.

TABLE II

BOUND COMPARISON FOR POLYHEDRAL UNCERTAINTY SETS

Problem	Average % Gap	% Tight Inst.	% Opt. Detections
$\mathscr{P}(Z_v)$	7.05	100	100
$\mathscr{P}(\Delta)$	10.27	49.75	99.6
$\mathscr{P}(\{\xi_w\})$	27.21	2.15	4.317
$\mathscr{L}(\mathbb{P}_{Z_v})$	39.39	0	0
$\mathscr{L}(\mathbb{P}_{\Delta})$	22.77	3.95	8.032
$\mathscr{L}(\mathbb{P}_{\Xi})$	45.61	0	0

TABLE III

BOUND COMPARISON FOR SPHERICAL UNCERTAINTY SETS

Problem:	$\mathscr{P}(\Delta)$	$\mathscr{P}(\{\xi_m\})$	$\mathscr{L}(\mathbb{P}_{\Delta})$	$\mathscr{L}(\mathbb{P}_{\Xi})$
% Gap:	18.76	34.43	37.17	47.73

We use this recipe to generate 1000 instances of $\mathscr{P}$ for each $p \in \{1,2,\infty\}$. The statistics for instances with box ($p=\infty$) and diamond ($p=1$) uncertainty sets are shown in Table II. In these cases, the vertices of Ξ can be enumerated. Thus, we can count how often any particular bound coincides with the true optimal value of $\mathscr{P}$ (tight instances), and how often the bound detects the optimality of LDRs among all instances for which LDRs are optimal (opt. detection). The results for the spherical uncertainty set ($p=2$) are shown in Table III.

APPENDIX

Proof of Theorem 4.4. We first derive an explicit formulation of problem $\mathscr{U}(\Delta)$. Let $[\Delta]$ be the matrix whose columns are the elements of the critical set $\Delta \subset \Xi$. Recall that $|\Delta| \le m+1$. Thus, the matrix $[\Delta]$ has at most $m+1$ columns of dimension $k+1$. We can write $\mathscr{U}(\Delta)$ as

$$\begin{array}{ll} \inf & c^\top x + t \\ \text{s.t.} & d^\top Y[\Delta] - t e_0^\top [\Delta] \le 0 \\ & Axe_0^\top [\Delta] + BY[\Delta] - C[\Delta] \le 0. \end{array} \tag{$\mathscr{U}(\Delta)$}$$

We construct the set Δ in a particular manner, such that any solution (x^*, Y^*, t^*) of $\tilde{\mathscr{U}}$ will also be optimal in $\mathscr{U}(\Delta)$, yielding the relation $\mathscr{U}^*(\Delta) = \tilde{\mathscr{U}}^* = \mathscr{U}^*$.

a) KKT conditions for $\tilde{\mathscr{U}}$: Consider the Lagrangian function of $\tilde{\mathscr{U}}$ described in (2), where $\lambda \in \mathbb{R}^{k+1}$ and $\Lambda \in \mathbb{R}^{(k+1)\times m}$ are the dual multipliers for the conic inequality constraints. From (2), we derive the KKT conditions of $\tilde{\mathscr{U}}$:

$$\begin{array}{rl} \text{Primal Feasibility:} & \left\{ \begin{array}{l} (d^\top Y - t e_0^\top)^\top \preceq_{\mathscr{K}^*} 0 \\ (Axe_0^\top + BY - C)^\top \preceq_{\mathscr{K}^*} 0 \end{array} \right. \\ \text{Dual Feasibility:} & \left\{ \begin{array}{l} 1 - e_0^\top \lambda = 0, \\ \Lambda B + \lambda d^\top = 0 \\ e_0^\top \Lambda A + c^\top = 0 \\ (\lambda, \Lambda) \succeq_{\mathscr{K}} 0 \end{array} \right. \\ \text{Complementarity:} & \left\{ \begin{array}{l} \left(d^\top Y - t e_0^\top\right)\lambda = 0 \\ \mathrm{tr}(\left(Axe_0^\top + BY - C\right)\Lambda) = 0. \end{array} \right. \end{array} \tag{8}$$

Let $(x^*, Y^*, t^*; \lambda^*, \Lambda^*)$ be a KKT point of $\tilde{\mathscr{U}}$. We can obtain (λ^*, Λ^*) by solving the dual of $\tilde{\mathscr{U}}$. From the KKT conditions we have that $(\lambda^*, \Lambda^*) \succeq_{\mathscr{K}} 0$ and $\lambda^* \in \Xi$. For each column $\Lambda^*_{(i)}$, $i = 1, \ldots, m$, there are two possibilities:

1) $\Lambda^*_{(i)} \ne 0$. As $\Lambda^*_{(i)} \ne 0$ and $\Lambda^*_{(i)} \in K$, then $e_0^\top \Lambda^*_{(i)} > 0$ (since Ξ is compact). Thus, there exists a positive scalar s_i so that $\xi_i = s_i \Lambda^*_{(i)} \in \Xi$. Note that ξ_i makes the i-th constraint binding, i.e., $A^{(i)}x^* + B^{(i)}Y^*\xi_i = C^{(i)}\xi_i$.
2) $\Lambda^*_{(i)} = 0$. As $\Lambda^*_{(i)}$ is vanishing, the i-th constraint does not bind the optimal LDR and can be omitted from (2) without affecting the remaining KKT point.

Furthermore, $\xi_0 = \lambda^* \in \Xi$ is a worst-case realization for the optimal LDR, meaning that $d^\top Y^* \xi_0 = t^*$. We can now construct the critical set as $\Delta = \{\xi_i : i \in I\} \bigcup \{\xi_0\}$, where I is the index set of the nonzero columns of Λ^*.

b) KKT conditions for $\mathscr{U}(\Delta)$: Let $v \in \mathbb{R}^{m+1}$ and $V \in \mathbb{R}^{(m+1)\times m}$ be the dual multipliers for the linear inequality constraints of $\mathscr{U}(\Delta)$. The KKT conditions of $\mathscr{U}(\Delta)$ are:

$$\begin{array}{rl} \text{Primal Feasibility:} & \left\{ \begin{array}{l} d^\top Y[\Delta] - t e_0^\top [\Delta] \le 0 \\ Axe_0^\top [\Delta] + BY[\Delta] - C[\Delta] \le 0 \end{array} \right. \\ \text{Dual Feasibility:} & \left\{ \begin{array}{l} 1 - e_0^\top [\Delta] v = 0 \\ [\Delta] VB + [\Delta] v d^\top = 0 \\ e_0^\top [\Delta] VA + c^\top = 0 \\ (v, V) \ge 0 \end{array} \right. \\ \text{Complementarity:} & \left\{ \begin{array}{l} \left(d^\top Y - t e_0^\top\right)[\Delta] v = 0 \\ \mathrm{tr}\left(\left(Axe_0^\top + BY - C\right)[\Delta] V\right) = 0. \end{array} \right. \end{array} \tag{9}$$

When Δ is defined as before, it is easy to show that there exists $(v^*, V^*) \ge 0$, so that $[\Delta] v^* = \lambda^*$ and $[\Delta] V^* = \Lambda^*$. As a result, the dual feasibility and complementarity conditions in (9) become identical to those in (8). Furthermore, any solution of $\tilde{\mathscr{U}}$ will satisfy the primal feasibility conditions in (9) since $\mathscr{U}(\Delta)$ is a relaxation of $\tilde{\mathscr{U}}$. Thus, the optimal solution (x^*, Y^*, t^*) of $\tilde{\mathscr{U}}$ is also optimal in $\mathscr{U}(\Delta)$. ■

REFERENCES

[1] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton University Press, 2009.

[2] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski, “Adjustable robust solutions of uncertain linear programs,” *Math. Program.*, vol. 99, no. 2, pp. 351–376, 2004.

[3] D. Bertsimas and V. Goyal, “On the power and limitations of affine policies in two-stage adaptive optimization,” *Math. Program.*, 2011, in press.

[4] J. Goh and M. Sim, “Distributionally robust optimization and its tractable approximations,” *Operat. Res.*, vol. 58, no. 4, pp. 902–917, 2010.

[5] X. Chen, M. Sim, P. Sun, and J. Zhang, “A linear decision-based approximation approach to stochastic programming,” *Operat. Res.*, vol. 56, no. 2, pp. 344–357, 2008.

[6] D. Bertsimas, D. Iancu, and P. Parrilo, “A hierarchy of near-optimal policies for multi-stage adaptive optimization,” *IEEE Trans. Automat. Contr.*, 2011, in press.

[7] ——, “Optimality of affine policies in multi-stage robust optimization,” *Math. of Operat. Res.*, vol. 35, no. 2, pp. 363–394, 2010.

[8] D. Kuhn, W. Wiesemann, and A. Georghiou, “Primal and dual linear decision rules in stochastic and robust optimization,” *Math. Program.*, pp. 1–33, 2009.

[9] A. Bemporad, F. Borrelli, and M. Morari, “Min-max control of constrained uncertain discrete-time linear systems,” *IEEE Trans. Automat. Contr.*, vol. 48, no. 9, pp. 1600–1606, 2003.

[10] A. Ben-Tal and A. Nemirovski, “Robust solutions of uncertain linear programs,” *Operat. Res. Letters*, vol. 25, no. 1, pp. 1–13, 1999.

[11] G. Calafiore and M. Campi, “Decision making in an uncertain environment: the scenario-based optimization approach.” *Math. Program.*, vol. 102, pp. 25–46, 2005.

[12] D. Bertsimas and C. Caramanis, “Adaptability via sampling,” in *46th IEEE Conf. on Decision and Control*, 2007, pp. 4717–4722.

[13] W. Wiesemann, D. Kuhn, and B. Rustem, “Robust resource allocations in temporal networks,” *Math. Program.*, 2011, in press.