Theory and applications of Robust Optimization

Dimitris Bertsimas[*], David B. Brown[†], Constantine Caramanis[‡]

First version: September 5, 2008
Revised version: October 27, 2010

Abstract

In this paper we survey the primary research, both theoretical and applied, in the area of Robust Optimization (RO). Our focus is on the computational attractiveness of RO approaches, as well as the modeling power and broad applicability of the methodology. In addition to surveying prominent theoretical results of RO, we also present some recent results linking RO to adaptable models for multi-stage decision-making problems. Finally, we highlight applications of RO across a wide spectrum of domains, including finance, statistics, learning, and various areas of engineering.

Keywords: Robust Optimization, robustness, adaptable optimization, applications of Robust Optimization.

[*]Sloan School of Management and Operations Research Center, Massachusetts Institute of Technology, E40-147, Cambridge, MA 02139. dbertsim@mit.edu

[†]Fuqua School of Business, Duke University, 100 Fuqua Drive, Box 90120, Durham, NC 27708. dbbrown@duke.edu

[‡]Department of Electrical and Computer Engineering, The University of Texas at Austin, 1 University Station, Austin, TX 78712. caramanis@mail.utexas.edu

1 Introduction

Optimization affected by parameter uncertainty has long been a focus of the mathematical programming community. Solutions to optimization problems can exhibit remarkable sensitivity to perturbations in the parameters of the problem (demonstrated in compelling fashion in [15]) thus often rendering a computed solution highly infeasible, suboptimal, or both (in short, potentially worthless).

In science and engineering, this is hardly a new notion. In the context of optimization, the most closely related field is that of Robust Control (we refer to the textbooks [136] and [67], and the references therein). While there are many high-level similarities, and indeed much of the motivation for the development of Robust Optimization came from the Robust Control community, Robust Optimization is a distinct field, focusing on traditional optimization-theoretic concepts, particularly algorithms, geometry, and tractability, in addition to modeling power and structural results which are more generically prevalent in the setting of robustness.

Stochastic Optimization starts by assuming the uncertainty has a probabilistic description. This approach has a long and active history dating at least as far back as Dantzig's original paper [61]. We refer the interested reader to several textbooks ([90, 39, 121, 92]) and the many references therein for a more comprehensive picture of Stochastic Optimization.

This paper considers Robust Optimization (RO), a more recent approach to optimization under uncertainty, in which the uncertainty model is not stochastic, but rather deterministic and set-based. Instead of seeking to immunize the solution in some probabilistic sense to stochastic uncertainty, here the decision-maker constructs a solution that is feasible for *any* realization of the uncertainty in a given set. The motivation for this approach is twofold. First, the model of set-based uncertainty is interesting in its own right, and in many applications is an appropriate notion of parameter uncertainty. Second, computational tractability is also a primary motivation and goal. It is this latter objective that has largely influenced the theoretical trajectory of Robust Optimization, and, more recently, has been responsible for its burgeoning success in a broad variety of application areas. The work of Ben-Tal and Nemirovski (e.g., [13, 14, 15]) and El Ghaoui et al. [77, 80] in the late 1990s, coupled with advances in computing technology and the development of fast, interior point methods for convex optimization, particularly for semidefinite optimization (e.g., Boyd and Vandenberghe, [42]) sparked a massive flurry of interest in the field of Robust Optimization.

Central issues we seek to address in this paper include tractability of robust optimization models; conservativeness of the RO formulation, and flexibility to apply the framework to different settings and applications. We give a summary of the main issues raised, and results presented.

1. Tractability: In general, the robust version of a tractable[1] optimization problem may not itself be tractable. We outline tractability results, which depend on the structure of the nominal problem as well as the class of uncertainty set. Many well-known classes of optimization problems, including LP, QCQP, SOCP, SDP, and some discrete problems as well, have a RO formulation that is tractable. Some care must be taken in the choice of the uncertainty set to ensure that tractability is preserved.

2. Conservativeness and probability guarantees: RO constructs solutions that are deterministically immune to realizations of the uncertain parameters in certain sets. This approach may be the only reasonable alternative when the parameter uncertainty is not stochastic, or if distributional information is not readily available. But even if there is an underlying distribution, the tractability benefits of the Robust Optimization approach may make it more attractive than alternative approaches from Stochastic Optimization. In this case, we might ask for probabilistic guarantees for the robust solution that can be computed *a priori*, as a function of the structure and size of the uncertainty set. In the sequel, we show that there are several convenient, efficient, and well-motivated parameterizations of different classes of uncertainty sets, that provide a notion of a *budget of uncertainty*. This allows the designer a level of flexibility in choosing the tradeoff between robustness and performance, and also allows the ability to choose the corresponding level of probabilistic protection. In particular, a perhaps surprising implication is that while the robust optimization formulation is inherently max-min (i.e., worst-case), the solutions it produces need not be overly conservative, and in many cases are very similar to those produced by stochastic methods.

3. Flexibility: In Section 2, we discuss a wide array of optimization classes, and also uncertainty sets, and consider the properties of the robust versions. In the final section of this paper, we illustrate the broad modeling power of Robust Optimization by presenting a wide variety of applications. We also give pointers to some surprising uses of robust optimization, particularly in statistics,

[1]Throughout this paper, we use the term "tractable" repeatedly. We use this as shorthand to refer to problems that can be reformulated into equivalent problems for which there are known solution algorithms with worst-case running time polynomial in a properly defined input size (see, e.g., Section 6.6 of Ben-Tal and Nemirovski [19]). Similarly, by "intractable" we mean the existence of such an algorithm for general instances of the problem would imply P=NP.

where robust optimization is used as a tool to imbue the solution with desirable properties, like sparsity, stability or statistical consistency.

The overall aim of this paper is to outline the development and main aspects of Robust Optimization, with an emphasis on its flexibility and structure. While the paper is organized around some of the main themes of robust optimization research, we attempt throughout to compare with other methods, particularly stochastic optimization, thus providing guidance and some intuition on when the robust optimization avenue may be most appropriate, and ultimately successful.

We also refer the interested reader to the recent book of Ben-Tal, El Ghaoui and Nemirovski [19], which is an excellent reference on Robust Optimization that provides more detail on specific formulation and tractability issues. Our goal here is to provide a more condensed, higher level summary of key methodological results as well as a broad array of applications that use Robust Optimization.

A First Example

To motivate RO and some of the modeling issues at hand, we begin with an example from portfolio selection. The example is a fairly standard one. We consider an investor who is attempting to allocate one unit of wealth among n risky assets with random return $\tilde{\boldsymbol{r}}$ and a risk-free asset (cash) with known return r_f. The investor may not short-sell risky assets or borrow. His goal is to optimally trade off between expected return and the probability that his portfolio loses money.

If the returns are stochastic with known distribution, the tradeoff between expected return and loss probability is a stochastic program. However, calculating a point on the pareto frontier is in general NP-hard even when the distribution of returns is discrete (Benati and Rizzi [20]).

We will consider two different cases: one where the distribution of asset price fluctuation matches the empirical distribution of given historical data and hence is known exactly, and then the case where it only approximately matches historical data. The latter case is of considerable practical importance, as the distribution of new returns (after an allocation decision) often deviate significantly from past samples. We compare the stochastic solution to several easily solved RO-based approximations in both of these cases.

The intractability of the stochastic problem arises because of the probability constraint on the loss:

$$\mathbb{P}(\tilde{\boldsymbol{r}}'\boldsymbol{x} + r_f(1 - \mathbf{1}'\boldsymbol{x}) \geq 1) \geq 1 - p_{loss}, \tag{1.1}$$

where $\boldsymbol{x}$ is the vector of allocations into the n risky assets (the decision variables). The robust optimization formulations replace this probabilistic constraint with a *deterministic constraint*, requiring the return to be nonnegative *for any realization of the returns in some given set*, called the uncertainty set:

$$\tilde{\boldsymbol{r}}'\boldsymbol{x} + r_f(1 - \mathbf{1}'\boldsymbol{x}) \geq 1 \ \forall \tilde{\boldsymbol{r}} \in \mathcal{R}. \tag{1.2}$$

While not explicitly specified in the robust constraint (1.2), the resulting solution has some p_{loss}. As a rough rule, the bigger the set $\mathcal{R}$, the lower the objective function (since there are more constraints to satisfy), and the smaller the loss probability p_{loss}. Central themes in robust optimization are understanding how to structure the uncertainty set $\mathcal{R}$ so that the resulting problem is tractable and favorably trades off expected return with loss probability p_{loss}. Section 2 is devoted to the tractability of different types of uncertainty sets. Section 3 focuses on obtaining *a priori* probabilistic guarantees given different uncertainty sets. Here, we consider three types of uncertainty sets, all defined with a parameter to control "size" so that we can explore the resulting tradeoff of return, and p_{loss}:

$$\begin{aligned}
\mathcal{R}^Q(\gamma) &= \left\{\tilde{\boldsymbol{r}} \ : \ (\tilde{\boldsymbol{r}} - \hat{\boldsymbol{r}})'\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{r}} - \hat{\boldsymbol{r}}) \leq \gamma^2\right\}, \\
\mathcal{R}^D(\Gamma) &= \left\{\tilde{\boldsymbol{r}} \ : \exists \boldsymbol{u} \in \mathbb{R}^n_+ \text{ s.t. } \tilde{r}_i = \hat{r}_i + (\underline{r}_i - \hat{r}_i)u_i, \ u_i \leq 1, \sum_{i=1}^n u_i \leq \Gamma\right\}, \\
\mathcal{R}^T(\alpha) &= \left\{\tilde{\boldsymbol{r}} \ : \ \exists \boldsymbol{q} \in \mathbb{R}^N_+ \text{ s.t. } \tilde{\boldsymbol{r}} = \sum_{i=1}^N q_i \boldsymbol{r}^i, \ \mathbf{1}'\boldsymbol{q} = 1, \ q_i \leq \frac{1}{N(1-\alpha)}, \ i = 1, \ldots, N\right\}.
\end{aligned}$$

The set $\mathcal{R}^Q(\gamma)$ is a quadratic or ellipsoidal uncertainty set: this set considers all returns within a radius of γ from the mean return vector, where the ellipsoid is tilted by the covariance. When $\gamma = 0$, this set is just the singleton $\{\hat{\boldsymbol{r}}\}$. The set $\mathcal{R}^D(\Gamma)$ (D for "D-norm" model considered in Section 2) considers all returns such that each component of the return is in the interval $[\underline{r}_i, \hat{r}_i]$, with the restriction that the total weight of deviation from $\hat{r}_i$, summed across all assets, may be no more than Γ. When $\Gamma = 0$, this set is the singleton $\{\hat{\boldsymbol{r}}\}$; at the other extreme, when $\Gamma = n$, returns in the range $[\underline{r}_i, \hat{r}_i]$ for all assets are considered. Finally, $\mathcal{R}^T(k)$ is the "tail" uncertainty set, and considers the convex hull of all possible $N(1-\alpha)$ point averages of the N returns. When $\alpha = 0$, this set is the singleton $\{\hat{\boldsymbol{r}}\}$. When $\alpha = (N-1)/N$, this set is the convex hull of all N returns.

To illustrate the use of these formulations, consider $n = 10$ risky assets based on $N = 300$ past market returns. The assets are a collection of equity and debt indices, and the return observations are monthly from a data set starting in 1981. For each of the three uncertainty RO formulations, we solve 200 problems, each maximizing expected return subject to feasibility and the robust constraint at one of 200 different values of their defining parameter γ, Γ, or α. In total, we solve 600 RO formulations. For comparison, we also formulate the problem of minimizing probability of loss subject to an expected return constraint as a stochastic program (which can be formulated as a mixed integer program), and

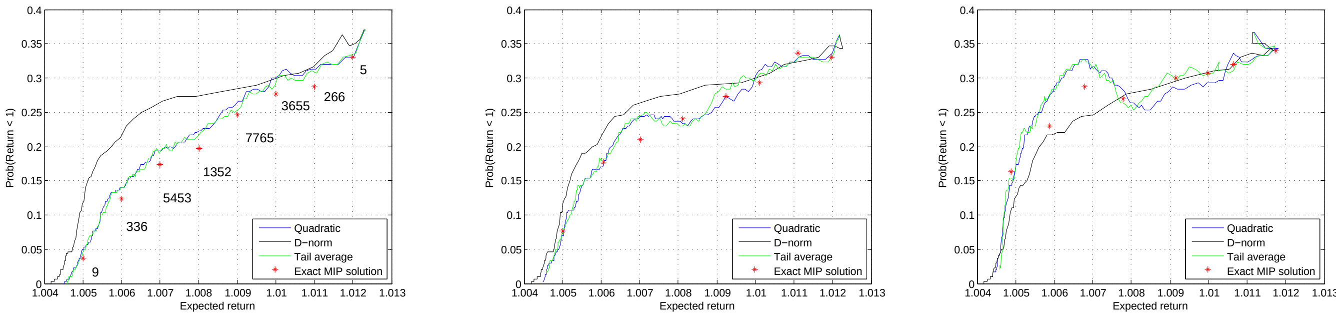

Figure 1: (L): Expected return-loss probability frontier for RO-based formulations and exact stochastic formulation; numbers are time (sec.) for solving each stochastic program. (C/R): Frontier for model with random perturbations bounded by 1% (C) and 2% (R).

solve 8 versions of this problem, each corresponding to one of 8 different expected return levels. The computations are performed using the MOSEK optimization toolbox in Matlab on a laptop computer with a 2.13GHZ processor and 2GB of RAM.

The results are shown in Figure 1. On the left, we see the frontier for the three RO-based formulations as well as the performance of the exact formulation (at the 8 return levels). The numbers indicate the time in seconds to solve the stochastic program in each case.

The stochastic model is designed for the nominal case, so we expect it to outperform the three RO-based formulations. However, even under this model, the gap from the $\mathcal{R}^Q$ and $\mathcal{R}^T$ RO frontiers is small: in several of the cases, the difference in performance is almost negligible. The largest improvement offered by the stochastic formulation is around a $2-3\%$ decrease in loss probability. Here, the solutions from the $\mathcal{R}^D$ model do not fare as well; though there is a range in which its performance is comparable to the other two RO-based models, typically its allocations appear to be conservative. In general, solving the stochastic formulation exactly is difficult, which is not surprising given its NP-hardness. Though a few of the instances at extreme return levels are solved in only a few seconds, several of the instances require well over an hour to solve, and the worst case requires over 2.1 hours to solve. The total time to solve these 8 instances is about 5.2 hours; by contrast, solving the 600 RO-based instances takes a bit under 10 minutes in total, or about one second per instance.

On the center and right parts of Figure 1 are results for the computed portfolios under the same return model but with random perturbations. Specifically, we perturb each of the $N \times n$ return values by a random number uniformly distributed on $[.99, 1.01]$ in the middle figure and $[.98, 1.02]$ in the right figure. At the 1% perturbation level, the gap in performance between the models is reduced, and there are regions in which each of the models are best as well as worst. The model based on $\mathcal{R}^D$ is least affected by the perturbation; its frontier is essentially unchanged. The models based on $\mathcal{R}^Q$ and $\mathcal{R}^T$ are more significantly affected, perhaps with the effect on $\mathcal{R}^T$ being a bit more pronounced. Finally, the stochastic formulation's solutions are the most sensitive of the bunch: though the SP solution is a winner in one of the 8 cases, it is worse off than the others in several of the other cases, and the increase in loss probability from the original model is as large as $5-6\%$ for the SP solutions.

At the 2% level, the results are even more pronounced: here, the SP solutions are always outperformed by one of the robust approaches, and the solutions based on $\mathcal{R}^D$ are relatively unaffected by the noise. The other two robust approaches are substantially affected, but nonetheless still win out in some parts of the frontier. When noise is introduced, it does not appear that the exact solutions confer much of an advantage and, in fact, may perform considerably worse. Though this is only one random trial, such results are typical.

There are several points of discussion here. First is the issue of complexity. The RO-based models are all fairly easy to solve here, though they themselves have complexities that scale differently. The $\mathcal{R}^Q$ model may be formulated as a second-order cone program (SOCP); both the $\mathcal{R}^D$ and the $\mathcal{R}^T$ models may be formulated as an LP. Meanwhile, the exact stochastic model is an NP-hard mixed integer program. Under the original model, it is clearly much easier to solve these RO-based models than the exact formulation. In a problem with financial data, it is easy to imagine having thousands of return samples. Whereas the RO formulations can still be solved quickly in such cases, solving the exact SP could be hopeless.

A second issue is the ability of solution methods to cope with deviations in the underlying model (or "model uncertainty"). The RO-based formulations themselves are different in this regard. Here, the $\mathcal{R}^D$ approach focuses on the worst-case returns on a subset of the assets, the $\mathcal{R}^Q$ approach focuses on the first two moments of the returns, and the $\mathcal{R}^T$ approach focuses on averages over the lower tail of the distribution. Though all of these are somehow "robust," $\mathcal{R}^D$ is the "most robust" of the three; indeed, we also implemented perturbations at the 5% level and found its frontier is relatively unchanged, while the other three frontiers are severely distorted. Intuitively, we would expect models that are more robust will fare better in situations with new or altered data; indeed, we will later touch upon some work that shows that there are intimate connections between the robustness of a model and its ability to generalize in a statistical learning sense.

This idea - that Robust Optimization is useful in dealing with erroneous or noise-corrupted data - seems relatively well understood by the optimization community (those who build, study, and solve optimization models) at-large. In fact, we would guess that many figure this to be the *raison d'être* for Robust Optimization. The final point that we would like to make is that, while dealing with perturbations is one virtue of the approach, RO is also more broadly of use as a computationally viable way to handle uncertainty in models that are *on their own* quite difficult to solve, as illustrated here.

In this example, even if we are absolutely set on the original model, it is hard to solve exactly. Nonetheless, two of the RO-based approaches perform well and are not far from optimal under the nominal model. In addition, they may be computed orders of magnitude faster than the exact solution. Of course, we also see that the user needs to have some understanding of the structure of the uncertainty set in order to intelligently use RO techniques: the approach with $\mathcal{R}^D$, though somewhat conservative in the original model, is quite resistant to perturbations of the model.

In short, RO provides a set of tools that may be useful in dealing with different types of uncertainties - both the "model error" or "noisy data" type as well as complex, stochastic descriptions of uncertainty in an explicit model - in a computationally manageable way. Like any approach, however, there are tradeoffs, both in terms of performance issues and in terms of problem complexity. Understanding and managing these tradeoffs requires expertise. The goal of this paper, first and foremost, is to describe some of this landscape for RO. This includes detailing what types of RO formulations may be efficiently solved in large scale, as well what connections various RO formulations have to perhaps more widely known methods. The second goal of this paper is to then highlight an array of application domains for which some of these techniques have been useful.

2 Structure and tractability results

In this section, we outline several of the structural properties, and detail some tractability results of Robust Optimization. We also show how the notion of a budget of uncertainty enters into several different uncertainty set formulations.

2.1 Robust Optimization

Given an objective $f_0(\boldsymbol{x})$ to optimize, subject to constraints $f_i(\boldsymbol{x}, \boldsymbol{u}_i) \leq 0$ with uncertain parameters, $\{\boldsymbol{u}_i\}$, the general Robust Optimization formulation is:

$$\begin{array}{ll} \text{minimize} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}, \boldsymbol{u}_i) \leq 0, \qquad \forall\, \boldsymbol{u}_i \in \mathcal{U}_i,\ i = 1, \ldots, m. \end{array} \tag{2.3}$$

Here $\boldsymbol{x} \in \mathbb{R}^n$ is a vector of decision variables, $f_0, f_i : \mathbb{R}^n \to \mathbb{R}$ are functions, and the uncertainty parameters $\boldsymbol{u}_i \in \mathbb{R}^k$ are assumed to take arbitrary values in the *uncertainty sets* $\mathcal{U}_i \subseteq \mathbb{R}^k$, which, for our purposes, will always be closed. The goal of (2.3) is to compute minimum cost solutions $\boldsymbol{x}^*$ among all those solutions which are feasible for *all* realizations of the disturbances $\boldsymbol{u}_i$ within $\mathcal{U}_i$. Thus, if some of the $\mathcal{U}_i$ are continuous sets, (2.3), as stated, has an infinite number of constraints. Intuitively, this problem offers some measure of feasibility protection for optimization problems containing parameters which are not known exactly.

It is worthwhile to notice the following, straightforward facts about the problem statement of (2.3):

- The fact that the objective function is unaffected by parameter uncertainty is without loss of generality; we may always introduce an auxiliary variable, call it t, and minimize t subject to the additional constraint $\max_{\boldsymbol{u}_0 \in \mathcal{U}_0} f_0(\boldsymbol{x}, \boldsymbol{u}_0) \leq t$.

- It is also without loss of generality to assume that the uncertainty set $\mathcal{U}$ has the form $\mathcal{U} = \mathcal{U}_1 \times \ldots \times \mathcal{U}_m$. If we have a single uncertainty set $\mathcal{U}$ for which we require $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m) \in \mathcal{U}$, then the constraint-wise feasibility requirement implies an equivalent problem is (2.3) with the $\mathcal{U}_i$ taken as the projection of $\mathcal{U}$ along the corresponding dimensions (see Ben-Tal and Nemirovski, [14] for more on this).

- Constraints without uncertainty are also captured in this framework by assuming the corresponding $\mathcal{U}_i$ to be singletons.

- Problem (2.3) also contains the instances when the decision or disturbance vectors are contained in more general vector spaces than $\mathbb{R}^n$ or $\mathbb{R}^k$ (e.g., $\mathbb{S}^n$ in the case of semidefinite optimization) with the definitions modified accordingly.

Robust Optimization is distinctly different than *sensitivity analysis*, which is typically applied as a post-optimization tool for quantifying the change in cost of the associated optimal solution with small perturbations in the underlying problem data. Here, our goal is to compute fixed solutions that ensure feasibility *independent of the data*. In other words, such solutions have *a priori* ensured feasibility when the problem parameters vary within the prescribed uncertainty set, which may be large. We refer the reader to some of the standard optimization literature (e.g., Bertsimas and Tsitsiklis, [37], Boyd and Vandenberghe, [43]) and works on perturbation theory (e.g., Freund, [75], Renegar, [123]) for more on sensitivity analysis.

It is not at all clear when (2.3) is efficiently solvable. One might imagine that the addition of robustness to a general optimization problem comes at the expense of significantly increased computational complexity. It turns out that this is true: the robust counterpart to an arbitrary convex optimization problem is in general intractable ([13]; some approximation results for robust convex problems with a conic structure are discussed in [35]). Despite this, there are many robust problems that may be handled in a tractable manner, and much of the literature has focused on specifying classes of functions f_i, coupled with the types of uncertainty sets $\mathcal{U}_i$, that yield tractable robust counterparts. If we define the robust feasible set to be

$$X(\mathcal{U}) \quad = \quad \{\boldsymbol{x} \mid f_i(\boldsymbol{x}, \boldsymbol{u}_i) \leq 0 \ \forall \ \boldsymbol{u}_i \in \mathcal{U}_i, \ i = 1, \ldots, m\}, \tag{2.4}$$

then for the most part,[2] tractability is tantamount to $X(\mathcal{U})$ being convex in $\boldsymbol{x}$, with an efficiently computable separation test. More precisely, in the next section we show that this is related to the structure of a particular subproblem. We now present an abridged taxonomy of some of the main results related to this issue.

2.2 Robust linear optimization

The robust counterpart of a linear optimization problem is written, without loss of generality, as

$$\begin{array}{ll} \text{minimize} & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \qquad \forall \ \boldsymbol{a_1} \in \mathcal{U}_1, \ldots, \boldsymbol{a_m} \in \mathcal{U}_m, \end{array} \tag{2.5}$$

where $\boldsymbol{a}_i$ represents the i^{th} row of the uncertain matrix $\boldsymbol{A}$, and takes values in the uncertainty set $\mathcal{U}_i \subseteq \mathbb{R}^n$. Then, $\boldsymbol{a}_i^\top \boldsymbol{x} \leq b_i$, $\forall \boldsymbol{a}_i \in \mathcal{U}_i$, if and only if $\max_{\{\boldsymbol{a}_i \in \mathcal{U}_i\}} \boldsymbol{a}_i^\top \boldsymbol{x} \leq b_i$, $\forall\, i$. We refer to this as the *subproblem* which must be solved. Ben-Tal and Nemirovski [14] show that the robust LP is essentially always tractable for most practical uncertainty sets of interest. Of course, the resulting robust problem may no longer be an LP. We now provide some more detailed examples.

Ellipsoidal Uncertainty: Ben-Tal and Nemirovski [14], as well as El Ghaoui et al. [77, 80], consider ellipsoidal uncertainty sets. Controlling the size of these ellipsoidal sets, as in the theorem below, has the interpretation of a budget of uncertainty that the decision-maker selects in order to easily trade off robustness and performance.

[2] i.e., subject to a Slater condition.

Theorem 1. *(Ben-Tal and Nemirovski, [14]) Let $\mathcal{U}$ be "ellipsoidal," i.e.,*

$$\mathcal{U} = U(\Pi, \boldsymbol{Q}) = \{\Pi(\boldsymbol{u}) \mid \|\boldsymbol{Q}\boldsymbol{u}\| \leq \rho\},$$

where $\boldsymbol{u} \to \Pi(\boldsymbol{u})$ is an affine embedding of $\mathbb{R}^L$ into $\mathbb{R}^{m\times n}$ and $\boldsymbol{Q} \in \mathbb{R}^{M\times L}$. Then Problem (2.5) is equivalent to a second-order cone program (SOCP). Explicitly, if we have the uncertain optimization

$$\begin{array}{ll} \textit{minimize} & \boldsymbol{c}^\top \boldsymbol{x} \\ \textit{subject to} & \boldsymbol{a}_i \boldsymbol{x} \leq b_i, \quad \forall \boldsymbol{a}_i \in \mathcal{U}_i, \quad \forall i = 1, \ldots, m, \end{array}$$

where the uncertainty set is given as:

$$\mathcal{U} = \{(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m) : \boldsymbol{a}_i = \boldsymbol{a}_i^0 + \Delta_i u_i, \; i = 1, \ldots, m, \quad ||u||_2 \leq \rho\},$$

($\boldsymbol{a}_i^0$ denotes the nominal value) then the robust counterpart is:

$$\begin{array}{ll} \textit{mininize} & \boldsymbol{c}^\top \boldsymbol{x} \\ \textit{subject to} & \boldsymbol{a}_i^0 \boldsymbol{x} \leq b_i - \rho||\Delta_i \boldsymbol{x}||_2, \quad \forall i = 1, \ldots, m. \end{array}$$

The intuition is as follows: for the case of ellipsoidal uncertainty, the subproblem $\max_{\{\boldsymbol{a}_i \in \mathcal{U}_i\}} \boldsymbol{a}_i^\top \boldsymbol{x} \leq b_i, \forall\, i$, is an optimization over a quadratic constraint. The dual, therefore, involves quadratic functions, which leads to the resulting SOCP.

Polyhedral Uncertainty: Polyhedral uncertainty can be viewed as a special case of ellipsoidal uncertainty [14]. When $\mathcal{U}$ is polyhedral, the subproblem becomes linear, and the robust counterpart is equivalent to a linear optimization problem. To illustrate this, consider the problem:

$$\begin{array}{rl} \min: & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.}: & \max_{\{\boldsymbol{D}_i \boldsymbol{a}_i \leq \boldsymbol{d}_i\}} \boldsymbol{a}_i^\top \boldsymbol{x} \leq b_i, \quad i = 1, \ldots, m. \end{array}$$

The dual of the subproblem (recall that $\boldsymbol{x}$ is not a variable of optimization in the inner max) becomes:

$$\left[\begin{array}{rl} \max: & \boldsymbol{a}_i^\top \boldsymbol{x} \\ \text{s.t.}: & \boldsymbol{D}_i \boldsymbol{a}_i \leq \boldsymbol{d}_i \end{array} \right] \longleftrightarrow \left[\begin{array}{rl} \min: & \boldsymbol{p}_i^\top \boldsymbol{d}_i \\ \text{s.t.}: & \boldsymbol{p}_i^\top \boldsymbol{D}_i = \boldsymbol{x} \\ & \boldsymbol{p}_i \geq 0. \end{array} \right]$$

and therefore the robust linear optimization now becomes:

$$\begin{array}{rl} \min: & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.}: & \boldsymbol{p}_i^\top \boldsymbol{d}_i \leq b_i, \quad i = 1, \ldots, m \\ & \boldsymbol{p}_i^\top \boldsymbol{D}_i = \boldsymbol{x}, \quad i = 1, \ldots, m \\ & \boldsymbol{p}_i \geq 0, \quad i = 1, \ldots, m. \end{array}$$

Thus the size of such problems grows polynomially in the size of the nominal problem and the dimensions of the uncertainty set.

Cardinality Constrained Uncertainty: Bertsimas and Sim ([34]) define a family of polyhedral uncertainty sets that encode a budget of uncertainty in terms of cardinality constraints: the number of parameters of the problem that are allowed to vary from their nominal values. The uncertainty set $\mathcal{R}^D$ from our introductory example, is an instance of this. More generally, given an uncertain matrix, $\boldsymbol{A} = (a_{ij})$, suppose each component a_{ij} lies in $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$. Rather than protect against the case when every parameter can deviate, as in the original model of Soyster ([128]), we allow at most Γ_i coefficients of row i to deviate. Thus the positive number Γ_i denotes the budget of uncertainty for the i^{th} constraint, and just as with the ellipsoidal sizing, controls the trade-off between the optimality of the solution, and its robustness to parameter perturbation.[3] Given values $\Gamma_1, \ldots, \Gamma_m$, the robust formulation becomes:

$$\begin{array}{rlr}
\min: & \boldsymbol{c}^\top \boldsymbol{x} & \\
\text{s.t.}: & \sum_j a_{ij} x_j + \max_{\{S_i \subseteq J_i \,:\, |S_i| = \Gamma_i\}} \sum_{j \in S_i} \hat{a}_{ij} y_j \leq b_i & 1 \leq i \leq m \\
& -y_j \leq x_j \leq y_j & 1 \leq j \leq n \\
& \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u} & \\
& \boldsymbol{y} \geq \boldsymbol{0}. &
\end{array} \tag{2.6}$$

Because of the set-selection in the inner maximization, this problem is nonconvex. However, one can show that the natural convex relaxation is exact. Thus, relaxing and taking the dual of the inner maximization problem, one can show that the above is equivalent to the following linear formulation, and therefore is tractable (and moreover is a linear optimization problem):

$$\begin{array}{rlr}
\max: & \boldsymbol{c}^\top \boldsymbol{x} & \\
\text{s.t.}: & \sum_j a_{ij} x_j + z_i \Gamma_i + \sum_j p_{ij} \leq b_i & \forall i \\
& z_i + p_{ij} \geq \hat{a}_{ij} y_j & \forall i, j \\
& -y_j \leq x_j \leq y_j & \forall j \\
& \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u} & \\
& \boldsymbol{p} \geq \boldsymbol{0} & \\
& \boldsymbol{y} \geq \boldsymbol{0}. &
\end{array}$$

Norm Uncertainty: Bertsimas et al. [31] show that robust linear optimization problems with uncer-

[3] For the full details see [34].

tainty sets described by more general norms lead to convex problems with constraints related to the dual norm. Here we use the notation $\text{vec}(\boldsymbol{A})$ to denote the vector formed by concatenating all of the rows of the matrix $\boldsymbol{A}$.

Theorem 2. *(Bertsimas et al., [31]) With the uncertainty set*

$$\mathcal{U} \;=\; \{\boldsymbol{A} \mid \|\boldsymbol{M}(\text{vec}(\boldsymbol{A}) - \text{vec}(\bar{\boldsymbol{A}}))\| \le \Delta\},$$

where $\boldsymbol{M}$ is an invertible matrix, $\bar{\boldsymbol{A}}$ is any constant matrix, and $\|\cdot\|$ is any norm, Problem (2.5) is equivalent to the problem

$$\begin{array}{ll} \text{minimize} & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} & \bar{\boldsymbol{A}}_i^\top \boldsymbol{x} + \Delta\|(\boldsymbol{M}^\top)^{-1}\boldsymbol{x}_i\|^* \le b_i, \qquad i = 1, \ldots, m, \end{array}$$

where $\boldsymbol{x}_i \in \mathbb{R}^{(m\cdot n)\times 1}$ is a vector that contains $\boldsymbol{x} \in \mathbb{R}^n$ in entries $(i-1)\cdot n + 1$ through $i \cdot n$ and 0 everywhere else, and $\|\cdot\|^*$ is the corresponding dual norm of $\|\cdot\|$.

Thus the norm-based model shown in Theorem 2 yields an equivalent problem with corresponding dual norm constraints. In particular, the l_1 and l_∞ norms result in linear optimization problems, and the l_2 norm results in a second-order cone problem.

In short, for many choices of the uncertainty set, robust linear optimization problems are tractable.

2.3 Robust quadratic optimization

Quadratically constrained quadratic programs (QCQP) have defining functions $f_i(\boldsymbol{x}, \boldsymbol{u}_i)$ of the form

$$f_i(\boldsymbol{x}, \boldsymbol{u}_i) \;=\; \|\boldsymbol{A}_i\boldsymbol{x}\|^2 + \boldsymbol{b}_i^\top \boldsymbol{x} + c_i.$$

Second order cone programs (SOCPs) have

$$f_i(\boldsymbol{x}, \boldsymbol{u}_i) \;=\; \|\boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{b}_i\| - \boldsymbol{c}_i^\top \boldsymbol{x} - d_i.$$

For both classes, if the uncertainty set $\mathcal{U}$ is a single ellipsoid (called *simple ellipsoidal uncertainty*) the robust counterpart is a semidefinite optimization problem (SDP). If $\mathcal{U}$ is polyhedral or the intersection of ellipsoids, the robust counterpart is NP-hard (Ben-Tal and Nemirovski, [13, 14, 18, 35]).

Following [18], we illustrate here only how to obtain the explicit reformulation of a robust quadratic constraint, subject to simple ellipsoidal uncertainty. Consider the quadratic constraint

$$\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} \le 2\boldsymbol{b}^\top \boldsymbol{x} + \boldsymbol{c}, \quad \forall(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}) \in \mathcal{U}, \tag{2.7}$$

where the uncertainty set $\mathcal{U}$ is an ellipsoid about a nominal point $(\boldsymbol{A}^0, \boldsymbol{b}^0, \boldsymbol{c}^0)$:

$$\mathcal{U} \triangleq \left\{ (\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}) := (\boldsymbol{A}^0, \boldsymbol{b}^0, \boldsymbol{c}^0) + \sum_{l=1}^{L} \boldsymbol{u}_l (\boldsymbol{A}^l, \boldsymbol{b}^l, \boldsymbol{c}^l) \ : \ ||\boldsymbol{u}||_2 \leq 1 \right\}.$$

As in the previous section, a vector $\boldsymbol{x}$ is feasible for the robust constraint (2.7) if and only if it is feasible for the constraint:

$$\begin{bmatrix} \max: & \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A} \boldsymbol{x} - 2\boldsymbol{b}^\top \boldsymbol{x} - \boldsymbol{c} \\ \text{s.t.}: & (\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}) \in \mathcal{U} \end{bmatrix} \leq 0.$$

This is the maximization of a convex quadratic objective (when the variable is the matrix $\boldsymbol{A}$, $\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A} \boldsymbol{x}$ is quadratic and convex in $\boldsymbol{A}$ since $\boldsymbol{x}\boldsymbol{x}^\top$ is always semidefinite) subject to a single quadratic constraint. It is well-known that while this problem is not convex (we are maximizing a convex quadratic) it nonetheless enjoys a hidden convexity property (for an early reference, see Brickman [44]) that allows it to be reformulated as a (convex) semidefinite optimization problem. This is related to the so-called S-lemma (or S-procedure) in control (e.g., Boyd et al. [40], Pólik and Terlaky [119]):

The S-lemma essentially gives the boundary between what we can solve exactly, and where solving the subproblem becomes difficult. If the uncertainty set is an intersection of ellipsoids or polyhedral, then exact solution of the subproblem is NP-hard.[4]

Taking the dual of the SDP resulting from the S-lemma, we have an exact, convex reformulation of the subproblem in the RO problem.

Theorem 3. *Given a vector $\boldsymbol{x}$, it is feasible to the robust constraint (2.7) if and only if there exists a scalar $\tau \in \mathbb{R}$ such that the following matrix inequality holds:*

$$\left(\begin{array}{c|ccc|c} c^0 + 2\boldsymbol{x}^\top \boldsymbol{b}^0 - \tau & \frac{1}{2}c^1 + \boldsymbol{x}^\top \boldsymbol{b}^1 & \cdots & c^L + \boldsymbol{x}^\top \boldsymbol{b}^L & (\boldsymbol{A}^0 \boldsymbol{x})^\top \\ \hline \frac{1}{2}c^1 + \boldsymbol{x}^\top \boldsymbol{b}^1 & \tau & & & (\boldsymbol{A}^1 \boldsymbol{x})^\top \\ \vdots & & \ddots & & \vdots \\ \frac{1}{2}c^L + \boldsymbol{x}^\top \boldsymbol{b}^L & & & \tau & (\boldsymbol{A}^L \boldsymbol{x})^\top \\ \hline \boldsymbol{A}^0 \boldsymbol{x} & \boldsymbol{A}^1 \boldsymbol{x} & \cdots & \boldsymbol{A}^L \boldsymbol{x} & I \end{array} \right) \succeq \boldsymbol{0}.$$

2.4 Robust Semidefinite Optimization

With ellipsoidal uncertainty sets, robust counterparts of semidefinite optimization problems are, in general, NP-hard (Ben-Tal and Nemirovski, [13], Ben-Tal et al. [8]). Similar negative results hold even in the case of polyhedral uncertainty sets (Nemirovski, [110]). One exception (Boyd et al. [40]) is when

[4] Nevertheless, there are some approximation results available: [18].

the uncertainty set is represented as *unstructured norm-bounded uncertainty.* Such uncertainty takes the form

$$\boldsymbol{A}_0(\boldsymbol{x}) + \boldsymbol{L}'(\boldsymbol{x})\boldsymbol{\zeta}\boldsymbol{R}(\boldsymbol{x}) + \boldsymbol{R}(\boldsymbol{x})\boldsymbol{\zeta}\boldsymbol{L}'(\boldsymbol{x}),$$

where $\boldsymbol{\zeta}$ is a matrix with norm satisfying $\|\boldsymbol{\zeta}\|_{2,2} \leq 1$, $\boldsymbol{L}$ and $\boldsymbol{R}$ are affine in the decision variables $\boldsymbol{x}$, and at least one of $\boldsymbol{L}$ or $\boldsymbol{R}$ is independent of $\boldsymbol{x}$.

In the general case, however, robust SDP is an intractable problem. Computing approximate solutions, i.e., solutions that are robust *feasible* but not robust *optimal* to robust semidefinite optimization problems has, as a consequence, received considerable attention (e.g., [80], [17, 16], and [35]). These methods provide bounds by developing inner approximations of the feasible set. The goodness of the approximation is based on a measure of how close the inner approximation is to the true feasible set. Precisely, the measure for this is:

$$\rho(\mathrm{AR}:\mathrm{R}) \quad = \quad \inf\left\{\rho \geq 1 \mid X(\mathrm{AR}) \supseteq X(\mathcal{U}(\rho))\right\},$$

where $X(\mathrm{AR})$ is the feasible set of the approximate robust problem and $X(\mathcal{U}(\rho))$ is the feasible set of the original robust SDP with the uncertainty set "inflated" by a factor of ρ. When the uncertainty set has "structured norm bounded" form, Ben-Tal and Nemirovski [17] develop an inner approximation such that $\rho(\mathrm{AR}:\mathrm{R}) \leq \pi\sqrt{\mu}/2$, where μ is the maximum rank of the matrices describing $\mathcal{U}$.

There has recently been additional work on Robust Semidefinite Optimization, for example exploiting sparsity [115], as well as in the area of control [78, 49].

2.5 Robust discrete optimization

Kouvelis and Yu [95] study robust models for some discrete optimization problems, and show that the robust counterparts to a number of polynomially solvable combinatorial problems are NP-hard. For instance, the problem of minimizing the maximum shortest path on a graph with only two scenarios for the cost vector can be shown to be an NP-hard problem [95].

Bertsimas and Sim [33], however, present a model for cost uncertainty in which each coefficient c_j is allowed to vary within the interval $[\bar{c}_j, \bar{c}_j + d_j]$, with no more than $\Gamma \geq 0$ coefficients allowed to vary. They then apply this model to a number of combinatorial problems, i.e., they attempt to solve

$$\begin{array}{ll} \text{minimize} & \bar{\boldsymbol{c}}^\top \boldsymbol{x} + \displaystyle\max_{\{S \ \mid \ S\subseteq N,\ |S|\leq\Gamma\}} \sum_{j\in S} d_j x_j \\ \text{subject to} & \boldsymbol{x} \in X, \end{array}$$

where $N = \{1, \ldots, n\}$ and X is a fixed set. Under this model for uncertainty, the robust version of a combinatorial problem may be solved by solving no more than $n + 1$ instances of the underlying, nominal problem. This result extends to approximation algorithms for combinatorial problems. For network flow problems, the above model can be applied and the robust solution can be computed by solving a logarithmic number of nominal, network flow problems.

Atamtürk [3] shows that, under an appropriate uncertainty model for the cost vector in a mixed 0-1 integer program, there is a tight, linear programming formulation of the robust problem with size polynomial in the size of a tight linear programming formulation for the nominal problem.

3 Choosing Uncertainty Sets

In addition to tractability, a central question in the Robust Optimization literature has been probability guarantees on feasibility under particular distributional assumptions for the disturbance vectors. Specifically, what does robust feasibility imply about probability of feasibility, i.e., what is the smallest ϵ we can find such that

$$\boldsymbol{x} \in X(\mathcal{U}) \Rightarrow \mathbb{P}\left(f_i(\boldsymbol{x}, \boldsymbol{u}_i) > 0\right) \leq \epsilon,$$

under (ideally mild) assumptions on a distribution for $\boldsymbol{u}_i$?

Such implications may be used as guidance for selection of a parameter representing the size of the uncertainty set. More generally, there are fundamental connections between distributional ambiguity, measures of risk, and uncertainty sets in robust optimization. In this section, we briefly discuss some of the connections in this vein.

3.1 Probability Guarantees

Probabilistic constraints, often called chance constraints in the literature, have a long history in stochastic optimization. Many approaches have been considered to address the computational challenges they pose ([121, 111]), including work using sampling to approximate the chance constraints [47, 45, 71].

One of the early discussions of probability guarantees in RO traces back to Ben-Tal and Nemirovski [15], who propose a robust model based on ellipsoids of radius Ω in the context of robust LP. Under this model, if the uncertain coefficients have bounded, symmetric support, they show that the corresponding robust feasible solutions must satisfy the constraint with high probability. Specifically, consider a linear constraint $\sum_j \tilde{a}_{ij} x_j \leq b_i$, where the coefficients $\tilde{a}_{ij}$ are uncertain and given by $\tilde{a}_{ij} = (1 + \epsilon \xi_{ij}) a_{ij}$, where

a_{ij} is a "nominal" value for the coefficient and $\{\xi_{ij}\}$ are zero mean, independent over j, and supported on $[-1, 1]$. Then a robust constraint of the form

$$\sum_j a_{ij}x_j + \epsilon\Omega\sqrt{\sum_j a_{ij}^2 x_j^2} \le b_i^+,$$

implies the robust solution satisfies the constraint with probability at least $1 - e^{-\Omega^2/2}$. This bound holds for any such distribution on the finite support.

In a similar spirit, Bertsimas and Sim [34] propose an uncertainty set of the form

$$\mathcal{U}_\Gamma = \left\{ \bar{\boldsymbol{A}} + \sum_{j \in J} z_j \hat{a}_j \;\middle|\; \|\boldsymbol{z}\|_\infty \le 1,\ \sum_{j \in J} \mathbf{1}(z_j) \le \Gamma \right\}, \tag{3.8}$$

for the coefficients $\boldsymbol{a}$ of an uncertain, linear constraint. Here, $\mathbf{1} : \mathbb{R} \to \mathbb{R}$ denotes the indicator function of y, i.e., $\mathbf{1}(y) = 0$ if and only if $y = 0$, $\bar{\boldsymbol{A}}$ is a vector of "nominal" values, $J \subseteq \{1, \ldots, n\}$ is an index set of uncertain coefficients, and $\Gamma \le |J|$ is an integer[5] reflecting the number of coefficients which are allowed to deviate from their nominal values. The dual formulation of this as a linear optimization is discussed in Section 2. The following then holds.

Theorem 4. *(Bertsimas and Sim [34]) Let $\boldsymbol{x}^*$ satisfy the constraint*

$$\max_{\boldsymbol{a} \in \mathcal{U}_\Gamma} \boldsymbol{a}^\top \boldsymbol{x}^* \le b,$$

where $\mathcal{U}_\Gamma$ is as in (3.8). If the random vector $\tilde{\boldsymbol{a}}$ has independent components with a_j distributed symmetrically on $[\bar{a}_j - \hat{a}_j, \bar{a}_j + \hat{a}_j]$ if $j \in J$ and $a_j = \bar{a}_j$ otherwise, then

$$\mathbb{P}\left(\tilde{\boldsymbol{a}}^\top \boldsymbol{x}^* > b\right) \le e^{-\frac{\Gamma^2}{2|J|}}.$$

In the case of linear optimization with only partial moment information (specifically, known mean and covariance), Bertsimas et al. [31] prove guarantees for the general norm uncertainty model used in Theorem 2. For instance, when $\|\cdot\|$ is the Euclidean norm, and $\boldsymbol{x}^*$ is feasible to the robust problem, Theorem 2 can be shown [31] to imply the guarantee

$$\mathbb{P}\left(\tilde{\boldsymbol{a}}^\top \boldsymbol{x}^* > b\right) \le \frac{1}{1 + \Delta^2},$$

where Δ is the radius of the uncertainty set, and the mean and covariance are used for $\bar{\boldsymbol{A}}$ and $\boldsymbol{M}$, respectively.

[5] The authors also consider Γ non-integer, but we omit this straightforward extension for notational convenience.

For more general robust conic optimization problems, results on probability guarantees are more elusive. Bertsimas and Sim are able to prove probability guarantees for their approximate robust solutions in [35]. In Chen et al. [56], more general deviation measures are considered that capture distributional skewness, leading to improved probability guarantees. Also of interest is the work of Paschalidis and Kang on probability guarantees and uncertainty set selection when the entire distribution is available [116].

3.2 Distributional Uncertainty

The issue of limited distributional information is central and has been the subject of considerable research in the decision theory literature. This work closely connects to robustness considerations and provides potential guidance and economic meaning to the choice of particular uncertainty sets.

Consider a function $u(\boldsymbol{x}, \xi)$ where ξ is a random parameter on some measure space $(\Omega, \mathcal{F})$. For the purposes of this discussion, let u be a concave, nondecreasing payoff function. In many situations, it may be unreasonable to expect the decision maker to have a full description of the distribution of ξ, but instead knows the distribution to be confined to some set of distributions $\mathcal{Q}$. Using a well-known duality result that traces back to at least the robust statistics literature (e.g., Huber [89]), one can establish that for any set $\mathcal{Q}$, there exists a convex, non-increasing, translation-invariant, positive homogeneous function μ on the induced space of random variables, such that

$$\inf_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}} \left[u(\boldsymbol{x}, \xi) \right] \; \geq \; 0 \quad \Leftrightarrow \quad \mu(u(\boldsymbol{x}, \xi)) \; \leq \; 0. \tag{3.9}$$

The function in this representation falls precisely into the class of *coherent risk measures* popularized by Artzner et al. [2]. These functions provide an economic interpretation in terms of a capital requirement: if X is a random variable (e.g., return), $\mu(X)$ represents the amount of money required to be added to X in order to make it "acceptable," given utility function u. The properties listed above are natural in a risk management setting: monotonicity states that one position that always pays off more than another should be deemed less risky; translation invariance means the addition of a sure amount to a position reduces the risk by precisely that amount; positive homogeneity means risks scale equally with the size of the stakes; and convexity means diversification among risky positions should be encouraged.

The above observation implies an immediate connection between these risk management tools, distributional ambiguity, and robust optimization. These connections have been explored in recent work on robust optimization. Natarajan et al. [109] investigate this connection with a focus on inferring risk measures from uncertainty sets.

Bertsimas and Brown [23] examine the question from the opposite perspective: namely, with risk preferences specified by a coherent risk measure, they examine the implications for uncertainty set structure in robust linear optimization problems. Due to the duality above, a risk constraint of the form $\mu(\tilde{\boldsymbol{a}}'\boldsymbol{x} - b) \leq 0$ on a linear constraint with an uncertain vector $\tilde{\boldsymbol{a}}$ can be equivalently expressed as

$$\boldsymbol{a}'\boldsymbol{x} \quad \geq \quad b \qquad \forall\, \boldsymbol{a} \in \mathcal{U},$$

where $\mathcal{U} = \text{conv}\left(\{\mathbb{E}_{\mathbb{Q}}\left[\boldsymbol{a}\right] \ : \ \mathbb{Q} \in \mathcal{Q}\}\right)$ and $\mathcal{Q}$ is the generating family for μ.

For a concrete application of this, one of most famous coherent risk measures is the *conditional value-at-risk* (CVaR), defined as

$$\mu(X) \quad \triangleq \quad \inf_{\nu \in \mathbb{R}} \left\{ \nu + \frac{1}{\alpha}\mathbb{E}\left[(-\nu - X)^{+}\right] \right\},$$

for any $\alpha \in (0, 1]$. For atomless distributions, CVaR is equivalent to the expected value of the random variable conditional on it being in its lower α quantile.

Consider the case when the uncertain vector $\tilde{\boldsymbol{a}}$ follows a discrete distribution with support $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ and corresponding probabilities $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N\}$. The generating family for CVaR in this case is $\mathcal{Q} = \{\boldsymbol{q} \in \Delta^N \ : \ q_i \leq p_i/\alpha\}$. This leads to the uncertainty set

$$\mathcal{U} \quad = \quad \text{conv}\left(\left\{ \frac{1}{\alpha}\sum_{i \in I} p_i \boldsymbol{a}_i + \left(1 - \frac{1}{\alpha}\sum_{i \in I} p_i\right)\boldsymbol{a}_j \ : \ I \subseteq \{1, \ldots, N\},\ j \in \{1, \ldots, N\} \setminus I,\ \sum_{i \in I} p_i \leq \alpha \right\} \right).$$

This set is a polytope, and therefore the robust optimization problem in this case may be reformulated as a linear program. When $p_i = 1/N$ and $\alpha = j/N$ for some $j \in \mathbb{Z}_+$, this has the interpretation of the convex hull of all j-point averages of $\mathcal{A}$.

Despite its popularity, CVaR represents only a special case of a much broader class of coherent risk measures that are *comonotone*. These risk measures satisfy the additional property that risky positions that "move together" in all states cannot be used to hedge one another. Extending a result from Dellacherie [65], Schmeidler [124] shows that the class of such risk measures is precisely the same as the set of functions representable as *Choquet integrals* (Choquet, [59]). Choquet integrals are the expectation under a set function that is non-additive and are a classical approach towards dealing with ambiguous distributions in decision theory. Bertsimas and Brown [24] discuss how one can form uncertainty sets in RO with these types of risk measures on discrete event spaces.

The use of a discrete probability space may be justified in situations when samples of the uncertainty are available. Delage and Ye [64] have proposed an approach to the distribution-robust problem

$$\text{minimize}_{\boldsymbol{x} \in X} \max_{f_\xi \in \mathcal{D}} \mathbb{E}_\xi\left[h(\boldsymbol{x}, \xi)\right],$$

where ξ is a random parameter with distribution f_ξ on some set of distributions $\mathcal{D}$ supported on a bounded set $\mathcal{S}$, h is convex in the decision variable $\boldsymbol{x}$, and X is a convex set. They consider sets of distributions $\mathcal{D}$ based on moment uncertainty with a particular focus on sets that have uncertainty in the mean and covariance of ξ. They then consider the problem when one has independent samples $\xi_1, \ldots, \xi_M$ and focus largely on the set

$$\mathcal{D}_1(\mathcal{S}, \hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0, \gamma_1, \gamma_2) \triangleq$$
$$\left\{\mathbb{P}\left(\xi \in S\right) = 1 \;:\; (\mathbb{E}\left[\xi\right] - \hat{\boldsymbol{\mu}}_0)' \hat{\boldsymbol{\Sigma}}_0^{-1} (\mathbb{E}\left[\xi\right] - \hat{\boldsymbol{\mu}}_0) \leq \gamma_1,\; \mathbb{E}\left[(\xi - \hat{\boldsymbol{\mu}}_0)'(\xi - \hat{\boldsymbol{\mu}}_0)\right] \preceq \gamma_2 \hat{\boldsymbol{\Sigma}}_0\right\}.$$

The above problem can be solved in polynomial time, and, with proper choices of γ_1, γ_2 and M, the resulting optimal value provides an upper bound on the expected cost with high probability. In the case of h as a piecewise linear, convex function, the resulting problem reduces to solving an SDP. This type of approach seems highly practical in settings (prevalent in many applications, e.g., finance) where samples are the only relevant information a decision maker has on the underlying distribution.

Related to distributional uncertainty is the work in [132]. Here, Xu, Caramanis and Mannor show that any robust optimization problem is equivalent to a distributionally robust problem. Using this equivalence to robust optimization, they show how robustness can guarantee consistency in sampled problems, even when the nominal sampled problem fails to be consistent.

More general types of robust optimization models have been explored, and such approaches draw further connections to research in decision theory. Ben-Tal et al. [6] propose an approach called *soft robust optimization* applicable in settings of distributional ambiguity. They modify the constraint (3.9) and consider the more general constraint

$$\inf_{\mathbb{Q} \in \mathcal{Q}(\epsilon)} \mathbb{E}_{\mathbb{Q}}\left[f(\boldsymbol{x}, \xi)\right] \;\; \geq \;\; -\epsilon \qquad \forall \epsilon \geq 0,$$

where $\{\mathcal{Q}(\epsilon)\}_{\epsilon \geq 0}$ is a set of sets of distributions, nondecreasing and convex on $\epsilon \geq 0$. This set of constraints considers different sized uncertainty sets with increasingly looser feasibility requirements as the uncertainty size grows; as such, it provides a potentially less conservative approach to RO than (3.9). This approach connects to the approach of *convex risk measures* (Föllmer and Schied, [74]), a generalization of the coherent risk measures mentioned above. Under a particular form for $\mathcal{Q}(\epsilon)$ based on relative entropy deviations, this model recovers the *multiplier preferences* of Hansen and Sargent [87], who develop their approach from robust control ideas in order to deal with model mis-specification in the decision making of economic agents (see also Maccheroni et al. [102] for a generalization known as *variational preferences*).

In short, there has been considerable work done in the domain of uncertainty set construction for RO. Some of this work focuses on the highly practical matter of implied probability guarantees under mild distributional assumptions or under a sufficiently large number of samples; other work draws connections to objects that have been axiomatized and developed in the decision theory literature over the past several decades.

4 Robust Adaptable Optimization

Thus far this paper has addressed optimization in the static, or one-shot case: the decision-maker considers a single-stage optimization problem affected by uncertainty. In this formulation, all the decisions are implemented simultaneously, and in particular, before any of the uncertainty is realized. In dynamic (or sequential) decision-making problems this single-shot assumption is restrictive and conservative. For example, in the inventory control example we discuss below, this would correspond to making all ordering decisions up front, without flexibility to adapt to changing demand patterns.

Sequential decision-making appears in a broad range of applications in many areas of engineering and beyond. There has been extensive work in optimal and robust control (e.g., the textbooks [67, 136], or the articles [72, 83, 85, 93], and references therein), and approximate and exact dynamic programming (e.g., see the textbooks [22, 21, 122]). In this section, we consider modeling approaches to incorporate sequential decision-making into the robust optimization framework.

4.1 Motivation and Background

In what follows, we refer to the *static* solution as the case where the $\boldsymbol{x}_i$ are all chosen at time 1 before any realizations of the uncertainty are revealed. The *dynamic* solution is the fully adaptable one, where $\boldsymbol{x}_i$ may have arbitrary functional dependence on past realizations of the uncertainty.

The question as to when adaptability has value is an interesting one that has received some attention. The papers by Dean, Goemans and Vondrák ([62, 81]) consider the value of adaptability in the context of stochastic optimization problems. They show there that for the stochastic knapsack problem, the value of adaptability is bounded: the value of the optimal adaptive solution is no more than a constant factor times the value of the optimal non-adaptive solution. In [27], Bertsimas and Goyal consider a two-stage mixed integer stochastic optimization problem with uncertainty in the right-hand-side. They show that a static robust solution approximates the fully-adaptable two-stage solution for the stochastic problem to within a factor of two, as long as the uncertainty set and the underlying measure, are both symmetric.

Despite the results for these cases, we would generally expect approaches that explicitly incorporate adaptivity to substantially outperform static approaches in multi-period problems. There are a number of approaches.

Receding Horizon

The most straightforward extension of the single-shot Robust Optimization formulation to that of sequential decision-making, is the so-called receding horizon approach. In this formulation, the static solution over all stages is computed, and the first-stage decision is implemented. At the next stage, the process is repeated. In the control literature this is known as open-loop feedback. While this approach is typically tractable, in many cases it may be far from optimal. In particular, because it is computed without any adaptability, the first stage decision may be overly conservative.

Stochastic Optimization

In Stochastic Optimization, the basic problem of interest is the so-called complete recourse problem (for the basic definitions and setup, see [39, 90, 121], and references therein). In this setup, the feasibility constraints of a single-stage Stochastic Optimization problem are relaxed and moved into the objective function by assuming that after the first-stage decisions are implemented and the uncertainty realized, the decision-maker has some recourse to ensure that the constraints are satisfied. The canonical example is in inventory control where in case of shortfall the decision-maker can buy inventory at a higher cost (possibly from a competitor) to meet demand. Then the problem becomes one of minimizing expected cost of the two-stage problem. If there is no complete recourse (i.e., not every first-stage decision can be completed to a feasible solution via second-stage actions) and furthermore the impact and cost of the second-stage actions are uncertain at the first stage, the problem becomes considerably more difficult. The feasibility constraint in particular is much more difficult to treat, since it cannot be entirely brought into the objective function.

When the uncertainty is assumed to take values in a finite set of small cardinality, the two-stage problem is tractable, and even for larger cardinality (but still finite) uncertainty sets (called scenarios), large-scale linear programming techniques such as Bender's decomposition can be employed to obtain a tractable formulation (see, e.g., [37]). In the case of incomplete recourse where feasibility is not guaranteed, robustness of the first-stage decision may require a very large number of scenarios in order to capture enough of the structure of the uncertainty. In the next section, we discuss a robust adaptable approach called Finite Adaptability that seeks to circumvent this issue.

Finally, even for small cardinality sets, the multi-stage complexity explodes in the number of stages ([125]). This is a central problem of multi-stage optimization, in both the robust and the stochastic formulations.

Dynamic Programming

Sequential decision-making under uncertainty has traditionally fallen under the purview of Dynamic Programming, where many exact and approximate techniques have been developed – we do not review this work here, but rather refer the reader to the books [22], [21], and [122]. The Dynamic Programming framework has been extended to the robust Dynamic Programming and robust MDP setting, where the payoffs and the dynamics are not exactly known, in Iyengar [91] and Nilim and El Ghaoui [113], and then also in Xu and Mannor [134]. Dynamic Programming yields tractable algorithms precisely when the Dynamic Programming recursion does not suffer from the curse of dimensionality. As the papers cited above make clear, this is a fragile property of any problem, and is particularly sensitive to the structure of the uncertainty. Indeed, the work in [91, 113, 134, 63] assumes a special property of the uncertainty set ("rectangularity") that effectively means that the decision-maker gains nothing by having future stage actions depend explicitly on past realizations of the uncertainty.

This section is devoted precisely to this problem: the dependence of future actions on past realizations of the uncertainty.

4.2 Tractability of Robust Adaptable Optimization

The uncertain multi-stage problem with deterministic set-based uncertainty, i.e., the robust multi-stage formulation, was first considered in [10]. There, the authors show that the two-stage linear problem with deterministic uncertainty is in general NP-hard. Consider the generic two-stage problem:

$$\begin{aligned} \min: &\quad \boldsymbol{c}^\top \boldsymbol{x}_1 \\ \text{s.t.}: &\quad \boldsymbol{A}_1(\boldsymbol{u})\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{u})\boldsymbol{x}_2(\boldsymbol{u}) \leq \boldsymbol{b}, \quad \forall \boldsymbol{u} \in \mathcal{U}. \end{aligned} \tag{4.10}$$

Here, $\boldsymbol{x}_2(\cdot)$ is an arbitrary function of $\boldsymbol{u}$. We can rewrite this explicitly in terms of the feasible set for the first stage decision:

$$\begin{aligned} \min: &\quad \boldsymbol{c}^\top \boldsymbol{x}_1 \\ \text{s.t.}: &\quad \boldsymbol{x}_1 \in \left\{ \boldsymbol{x}_1 \,:\, \forall \boldsymbol{u} \in \mathcal{U},\, \exists \boldsymbol{x}_2 \text{ s.t. } \boldsymbol{A}_1(\boldsymbol{u})\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{u})\boldsymbol{x}_2 \leq \boldsymbol{b} \right\}. \end{aligned} \tag{4.11}$$

The feasible set is convex, but nevertheless the optimization problem is in general intractable. Consider a simple example given in [10]:

$$\begin{aligned} \min : \quad & x_1 \\ \text{s.t.} : \quad & x_1 - \boldsymbol{u}^\top \boldsymbol{x}_2(\boldsymbol{u}) \geq 0 \\ & \boldsymbol{x}_2(\boldsymbol{u}) \geq \boldsymbol{B}\boldsymbol{u} \\ & \boldsymbol{x}_2(\boldsymbol{u}) \leq \boldsymbol{B}\boldsymbol{u}. \end{aligned} \tag{4.12}$$

It is not hard to see that the feasible first-stage decisions are given by the set:

$$\{x_1 \,:\, x_1 \geq \boldsymbol{u}^\top \boldsymbol{B}\boldsymbol{u}, \quad \forall \boldsymbol{u} \in \mathcal{U}\}.$$

The set is, therefore, a ray in $\mathbb{R}^1$, but determining the left endpoint of this line requires computing a maximization of a (possibly indefinite) quadratic $\boldsymbol{u}^\top \boldsymbol{B}\boldsymbol{u}$, over the set $\mathcal{U}$. In general, this problem is NP-hard (see, e.g., [76]).

4.3 Theoretical Results

Despite the hardness result above, there has been considerable effort devoted to obtaining different approximations and approaches to the multi-stage optimization problem.

4.3.1 Affine Adaptability

In [10], the authors formulate an approximation to the general robust multi-stage optimization problem, which they call the *Affinely Adjustable Robust Counterpart* (AARC). Here, they explicitly parameterize the future stage decisions as affine functions of the revealed uncertainty. For the two-stage problem (4.10), the second stage variable, $\boldsymbol{x}_2(\boldsymbol{u})$, is parameterized as:

$$\boldsymbol{x}_2(\boldsymbol{u}) = \boldsymbol{Q}\boldsymbol{u} + \boldsymbol{q}.$$

Now, the problem becomes:

$$\begin{aligned} \min : \quad & \boldsymbol{c}^\top \boldsymbol{x}_1 \\ \text{s.t.} : \quad & \boldsymbol{A}_1(\boldsymbol{u})\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{u})[\boldsymbol{Q}\boldsymbol{u} + \boldsymbol{q}] \leq \boldsymbol{b}, \quad \forall \boldsymbol{u} \in \mathcal{U}. \end{aligned}$$

This is a single-stage RO. The decision-variables are $(\boldsymbol{x}_1, \boldsymbol{Q}, \boldsymbol{q})$, and they are all to be decided before the uncertain parameter, $\boldsymbol{u} \in \mathcal{U}$, is realized.

In the generic formulation of the two-stage problem (4.10), the functional dependence of $\boldsymbol{x}_2(\cdot)$ on $\boldsymbol{u}$ is arbitrary. In the affine formulation, the resulting problem is a linear optimization problem with uncertainty. The parameters of the problem, however, now have a quadratic dependence on the uncertain parameter $\boldsymbol{u}$. Thus in general, the resulting robust linear optimization will not be tractable – consider again the example (4.12).

Despite this negative result, there are some positive complexity results concerning the affine model. In order to present these, let us explicitly denote the dependence of the optimization parameters, $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, as:

$$[\boldsymbol{A}_1, \boldsymbol{A}_2](\boldsymbol{u}) = [\boldsymbol{A}_1^{(0)}, \boldsymbol{A}_2^{(0)}] + \sum_{l=1}^{L} u_l [\boldsymbol{A}_1^{(l)}, \boldsymbol{A}_2^{(l)}].$$

When we have $\boldsymbol{A}_2^{(l)} = \boldsymbol{0}$, for all $l \geq 1$, the matrix multiplying the second stage variables is constant. This setting is known as the case of *fixed recourse*. We can now write the second stage variables explicitly in terms of the columns of the matrix $\boldsymbol{Q}$. Letting $\boldsymbol{q}^{(l)}$ denote the l^{th} column of $\boldsymbol{Q}$, and $\boldsymbol{q}^{(0)} = \boldsymbol{q}$ the constant vector, we have:

$$\begin{aligned} \boldsymbol{x}_2 &= \boldsymbol{Q}\boldsymbol{u} + \boldsymbol{q}_0 \\ &= \boldsymbol{q}^{(0)} + \sum_{l=1}^{L} u_l \boldsymbol{q}^{(l)}. \end{aligned}$$

Letting $\boldsymbol{\chi} = (\boldsymbol{x}_1, \boldsymbol{q}^{(0)}, \boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(L)})$ denote the full decision vector, we can write the i^{th} constraint as

$$\begin{aligned} 0 &\leq (\boldsymbol{A}_1^{(0)} \boldsymbol{x}_1 + \boldsymbol{A}_2^{(0)} \boldsymbol{q}^{(0)} - \boldsymbol{b})_i + \sum_{l=1}^{L} u_l (\boldsymbol{A}_1^{(l)} \boldsymbol{x}_1 + \boldsymbol{A}_2 \boldsymbol{q}^{(l)})_i \\ &= \sum_{l=0}^{L} a_l^i(\boldsymbol{\chi}), \end{aligned}$$

where we have defined

$$a_l^i \triangleq a_l^i(\boldsymbol{\chi}) \triangleq (\boldsymbol{A}_1^{(l)} \boldsymbol{x}_1 + \boldsymbol{A}_2^{(l)} \boldsymbol{q}^{(l)})_i, \qquad a_0^i \triangleq (\boldsymbol{A}_1^{(0)} \boldsymbol{x}_1 + \boldsymbol{A}_2^{(0)} \boldsymbol{q}^{(0)} - \boldsymbol{b})_i.$$

Theorem 5 ([10])**.** *Assume we have a two-stage linear optimization with fixed recourse, and with conic uncertainty set:*

$$\mathcal{U} = \{\boldsymbol{u} \,:\, \exists \xi \text{ s.t. } \boldsymbol{V}_1 \boldsymbol{u} + \boldsymbol{V}_2 \boldsymbol{\xi} \geq_{\mathcal{K}} \boldsymbol{d}\} \subseteq \mathbb{R}^L,$$

where $\mathcal{K}$ is a convex cone with dual $\mathcal{K}^*$. If $\mathcal{U}$ has nonempty interior, then the AARC can be reformulated *as the following optimization problem:*

$$\begin{aligned} \min: \quad & \boldsymbol{c}^\top \boldsymbol{x}_1 \\ \text{s.t.}: \quad & \boldsymbol{V}_1 \lambda^i - a^i(\boldsymbol{x}_1, \boldsymbol{q}^{(0)}, \ldots, \boldsymbol{q}^{(L)}) = 0, \quad i = 1, \ldots, m \\ & \boldsymbol{V}_2 \lambda^i = 0, \quad i = 1, \ldots, m \\ & \boldsymbol{d}^\top \lambda^i + a_0^i(\boldsymbol{x}_1, \boldsymbol{q}^{(0)}, \ldots, \boldsymbol{q}^{(L)}) \geq 0, \quad i = 1, \ldots, m \\ & \lambda^i \geq_{\mathcal{K}^*} 0, \quad i = 1, \ldots, m. \end{aligned}$$

If the cone $\mathcal{K}$ is the positive orthant, then the AARC given above is an LP.

The case of non-fixed recourse is more difficult because of the quadratic dependence on $\boldsymbol{u}$. Note that the example in (4.12) above involves an uncertainty-affected recourse matrix. In this non-fixed recourse case, the robust constraints have a component that is quadratic in the uncertain parameters, $\boldsymbol{u}_i$. These robust constraints then become:

$$\left[\boldsymbol{A}_1^{(0)} + \sum \boldsymbol{u}_l A_1^{(1)}\right] \boldsymbol{x}_1 + \left[\boldsymbol{A}_2^{(0)} + \sum \boldsymbol{u}_l A_2^{(1)}\right] \left[\boldsymbol{q}^{(0)} + \sum \boldsymbol{u}_l \boldsymbol{q}^{(l)}\right] - \boldsymbol{b} \leq \boldsymbol{0}, \quad \forall \boldsymbol{u} \in \mathcal{U},$$

which can be rewritten to emphasize the quadratic dependence on $\boldsymbol{u}$, as

$$\left[\boldsymbol{A}_1^{(0)} \boldsymbol{x}_1 + \boldsymbol{A}_2^{(0)} \boldsymbol{q}^{(0)} - \boldsymbol{b}\right] + \sum \boldsymbol{u}_l \left[\boldsymbol{A}_1^{(l)} \boldsymbol{x}_1 + \boldsymbol{A}_2^{(0)} \boldsymbol{q}^{(l)} + \boldsymbol{A}_2^{(l)} \boldsymbol{q}^{(0)}\right] + \left[\sum \boldsymbol{u}_k \boldsymbol{u}_l \boldsymbol{A}_2^{(k)} \boldsymbol{q}^{(l)}\right] \leq 0, \quad \forall \boldsymbol{u} \in \mathcal{U}.$$

Writing

$$\begin{aligned} \boldsymbol{\chi} & \triangleq (\boldsymbol{x}_1, \boldsymbol{q}^{(0)}, \ldots, \boldsymbol{q}^{(L)}), \\ \alpha_i(\boldsymbol{\chi}) & \triangleq -[\boldsymbol{A}_1^{(0)} \boldsymbol{x}_1 + \boldsymbol{A}_2^{(0)} \boldsymbol{q}^{(0)} - \boldsymbol{b}]_i \\ \beta_i^{(l)}(\boldsymbol{\chi}) & \triangleq -\frac{[\boldsymbol{A}_1^{(l)} \boldsymbol{x}_1 + \boldsymbol{A}_2^{(0)} \boldsymbol{q}^{(l)} - \boldsymbol{b}]_i}{2}, \quad l = 1, \ldots, L \\ \Gamma_i^{(l,k)}(\boldsymbol{\chi}) & \triangleq -\frac{[\boldsymbol{A}_2^{(k)} \boldsymbol{q}^{(l)} + \boldsymbol{A}_2^{(l)} \boldsymbol{q}^{(k)}]_i}{2}, \quad l, k = 1, \ldots, L, \end{aligned}$$

the robust constraints can now be expressed as:

$$\alpha_i(\boldsymbol{\chi}) + 2\boldsymbol{u}^\top \beta_i(\boldsymbol{\chi}) + \boldsymbol{u}^\top \Gamma_i(\boldsymbol{\chi}) \boldsymbol{u} \geq 0, \quad \forall \boldsymbol{u} \in \mathcal{U}. \tag{4.13}$$

Theorem 6 ([10])**.** *Let our uncertainty set be given as the intersection of ellipsoids:*

$$\mathcal{U} \triangleq \{\boldsymbol{u} : \boldsymbol{u}^\top (\rho^{-2} S_k) \boldsymbol{u} \leq 1, \ k = 1, \ldots, K\},$$

where ρ controls the size of the ellipsoids. Then the original AARC problem can be approximated by the following semidefinite optimization problem:

$$
\begin{aligned}
\min : \quad & \boldsymbol{c}^\top \boldsymbol{x}_1 \\
\text{s.t.} : \quad & \left(\begin{array}{c|c} \Gamma_i(\boldsymbol{\chi}) + \rho^{-2} \sum_{k=1}^K \lambda_k S_k & \beta_i(\boldsymbol{\chi}) \\ \hline \beta_i(\boldsymbol{\chi})^\top & \alpha_i(\boldsymbol{\chi}) - \sum_{k=1}^K \lambda_k^{(i)} \end{array} \right) \succeq \mathbf{0}, \; i = 1, \ldots, m \\
& \lambda^{(i)} \geq 0, \; i = 1, \ldots, m
\end{aligned}
\tag{4.14}
$$

The constant ρ in the definition of the uncertainty set $\mathcal{U}$ can be regarded as a measure of the level of the uncertainty. This allows us to give a bound on the tightness of the approximation. Define the constant

$$\gamma \triangleq \sqrt{2 \ln \left(6 \sum_{k=1}^K \text{Rank}(S_k) \right)}.$$

Then we have the following.

Theorem 7 ([10])**.** *Let $\mathcal{X}_\rho$ denote the feasible set of the AARC with noise level ρ. Let $\mathcal{X}_\rho^{approx}$ denote the feasible set of the SDP approximation to the AARC with uncertainty parameter ρ. Then, for γ defined as above, we have the containment:*

$$\mathcal{X}_{\gamma\rho} \subseteq \mathcal{X}_\rho^{approx} \subseteq \mathcal{X}_\rho.$$

This tightness result has been improved; see [66].

There have been a number of applications building upon affine adaptability, in a wide array of areas:

1. Integrated circuit design: In [104], the affine adjustable approach is used to model the yield-loss optimization in chip design, where the first stage decisions are the pre-silicon design decisions, while the second-stage decisions represent post-silicon tuning, made after the manufacturing variability is realized and can then be measured.

2. Comprehensive Robust Optimization: In [7], the authors extend the robust static, as well as the affine adaptability framework, to soften the hard constraints of the optimization, and hence to reduce the conservativeness of robustness. At the same time, this controls the infeasibility of the solution even when the uncertainty is realized outside a nominal compact set. This has many applications, including portfolio management, and optimal control.

3. Network flows and Traffic Management: In [112], the authors consider the robust capacity expansion of a network flow problem that faces uncertainty in the demand, and also the travel time

along the links. They use the adjustable framework of [10], and they show that for the structure of uncertainty sets they consider, the resulting problem is tractable. In [107], the authors consider a similar problem under transportation cost and demand uncertainty, extending the work in [112].

4. Chance constraints: In [57], the authors apply a modified model of affine adaptability to the stochastic programming setting, and show how this can improve approximations of chance constraints. In [70], the authors formulate and propose an algorithm for the problem of two-stage convex chance constraints when the underlying distribution has some uncertainty (i.e., an *ambiguous* distribution).

5. Numerous other applications have been considered, including portfolio management [46, 129], coordination in wireless networks [135], robust control [84], and model adaptive control.

Additional work in affine adaptability has been done in [57], where the authors consider modified linear decision rules in the context of only partial distibutional knowledge, and within that framework derive tractable approximations to the resulting robust problems. See also Ben-Tal et al. [19] for a detailed discussion of affine decision rules in multistage optimization. Recently, [29] have given conditions under which affine policies are in fact optimal, and affine policies have been extended to higher order polynomial adaptability in [25, 28].

4.3.2 Finite Adaptability

The framework of Finite Adaptability, introduced in Bertsimas and Caramanis [26] and Caramanis [52], treats the discrete setting by modeling the second-stage variables, $\boldsymbol{x}_2(\boldsymbol{u})$, as piecewise constant functions of the uncertainty, with k pieces. One advantage of such an approach is that, due to the inherent finiteness of the framework, the resulting formulation can accommodate discrete variables. In addition, the level of adaptability can be adjusted by changing the number of pieces in the piecewise constant second stage variables. (For an example from circuit design where such second stage limited adaptability constraints are physically motivated by design considerations, see [103, 127]).

If the partition of the uncertainty set is fixed, then the resulting problem retains the structure of the original nominal problem, and the number of second stage variables grows by a factor of k. In general, computing the optimal partition into even two regions is NP-hard [26], however, if any one of the three quantities: (a) dimension of the uncertainty; (b) dimension of the decision-space; or (c) number of uncertain constraints, is small, then computing the optimal 2-piecewise constant second stage policy can be done efficiently. One application where the dimension of the uncertainty is large, but can be approximated by a low-dimensional set, is weather uncertainty in air traffic flow management (see [26]).

4.3.3 Network Design

In Atamturk and Zhang [4], the authors consider two-stage robust network flow and design, where the demand vector is uncertain. This work deals with computing the optimal second stage adaptability, and characterizing the first-stage feasible set of decisions. While this set is convex, solving the separation problem, and hence optimizing over it, can be NP-hard, even for the two-stage network flow problem.

Given a directed graph $G = (V, E)$, and a demand vector $\boldsymbol{d} \in \mathbb{R}^V$, where the edges are partitioned into first-stage and second-stage decisions, $E = E_1 \cup E_2$, we want to obtain an expression for the feasible first-stage decisions. We define some notation first. Given a set of nodes, $S \subseteq V$, let $\delta^+(S), \delta^-(S)$, denote the set of arcs into and out of the set S, respectively. Then, denote the set of flows on the graph satisfying the demand by

$$\mathcal{P}_{\boldsymbol{d}} \stackrel{\triangle}{=} \{\boldsymbol{x} \in \mathbb{R}_+^E \,:\, \boldsymbol{x}(\delta^+(i)) - \boldsymbol{x}(\delta^-(i)) \geq d_i, \ \forall i \in V\}.$$

If the demand vector $\boldsymbol{d}$ is only known to lie in a given compact set $\mathcal{U} \subseteq \mathbb{R}^V$, then the set of flows satisfying every possible demand vector is given by the intersection $\mathcal{P} = \bigcap_{\boldsymbol{d}\in\mathcal{U}} \mathcal{P}_{\boldsymbol{d}}$. If the edge set E is partitioned $E = E_1 \cup E_2$ into first and second-stage flow variables, then the set of first-stage-feasible vectors is:

$$\mathcal{P}(E_1) \stackrel{\triangle}{=} \bigcap_{\boldsymbol{d}\in\mathcal{U}} \mathrm{Proj}_{E_1} \mathcal{P}_{\boldsymbol{d}},$$

where $\mathrm{Proj}_{E_1}\mathcal{P}_{\boldsymbol{d}} \stackrel{\triangle}{=} \{\boldsymbol{x}_{E_1} \,:\, (\boldsymbol{x}_{E_1}, \boldsymbol{x}_{E_2}) \in \mathcal{P}_{\boldsymbol{d}}\}$. Then we have:

Theorem 8 ([4])**.** *A vector* $\boldsymbol{x}_{E_1}$ *is an element of* $\mathcal{P}(E_1)$ *iff* $\boldsymbol{x}_{E_1}(\delta^+(S)) - \boldsymbol{x}_{E_1}(\delta^-(S)) \geq \zeta_S$*, for all subsets* $S \subseteq V$ *such that* $\delta^+(S) \subseteq E_1$*, where we have defined* $\zeta_S \stackrel{\triangle}{=} \max\{\boldsymbol{d}(S) \,:\, \boldsymbol{d} \in \mathcal{U}\}$*.*

The authors then show that for both the budget-restricted uncertainty model, $\mathcal{U} = \{\boldsymbol{d} \,:\, \sum_{i\in V} \pi_i d_i \leq \pi_0,\ \bar{\boldsymbol{d}} - \boldsymbol{h} \leq \boldsymbol{d} \leq \bar{\boldsymbol{d}} + \boldsymbol{h}\}$, and the cardinality-restricted uncertainty model, $\mathcal{U} = \{\boldsymbol{d} \,:\, \sum_{i\in V} \lceil |d_i - \bar{d}_i| \backslash h_i \rceil \leq \Gamma,\ \bar{\boldsymbol{d}} - \boldsymbol{h} \leq \boldsymbol{d} \leq \bar{\boldsymbol{d}} + \boldsymbol{h}\}$, the separation problem for the set $\mathcal{P}(E_1)$ is NP-hard:

Theorem 9 ([4])**.** *For both classes of uncertainty sets given above, the separation problem for* $\mathcal{P}(E_1)$ *is NP-hard for bipartite* $G(V, B)$*.*

These results extend also to the framework of two-stage network design problems, where the capacities of the edges are also part of the optimization. If the second stage network topology is totally ordered, or an arborescence, then the separation problem becomes tractable.

5 Applications of Robust Optimization

In this section, we examine several applications approached by Robust Optimization techniques.

5.1 Portfolio optimization

One of the central problems in finance is how to allocate monetary resources across risky assets. This problem has received considerable attention from the Robust Optimization community and a wide array of models for robustness have been explored in the literature.

5.1.1 Uncertainty models for return mean and covariance

The classical work of Markowitz ([105, 106]) served as the genesis for modern portfolio theory. The canonical problem is to allocate wealth across n risky assets with mean returns $\boldsymbol{\mu} \in \mathbb{R}^n$ and return covariance matrix $\boldsymbol{\Sigma} \in \mathbb{S}^n_{++}$ over a weight vector $\boldsymbol{w} \in \mathbb{R}^n$. Two versions of the problem arise; first, the *minimum variance problem*, i.e.,

$$\min \left\{ \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \ : \ \boldsymbol{\mu}^\top \boldsymbol{w} \geq r, \ \boldsymbol{w} \in \mathcal{W} \right\}, \tag{5.15}$$

or, alternatively, the *maximum return problem*, i.e.,

$$\max \left\{ \boldsymbol{\mu}^\top \boldsymbol{w} \ : \ \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \leq \sigma^2, \boldsymbol{w} \in \mathcal{W} \right\}. \tag{5.16}$$

Here, r and σ are investor-specified constants, and $\mathcal{W}$ represents the set of acceptable weight vectors ($\mathcal{W}$ typically contains the normalization constraint $\boldsymbol{e}^\top \boldsymbol{w} = 1$ and often has "no short-sales" constraints, i.e., $w_i \geq 0, \ i = 1, \ldots, n$, among others).

While this framework proposed by Markowitz revolutionized the financial world, particularly for the resulting insights in trading off *risk* (variance) and *return*, a fundamental drawback from the practitioner's perspective is that $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are rarely known with complete precision. In turn, optimization algorithms tend to exacerbate this problem by finding solutions that are "extreme" allocations and, in turn, very sensitive to small perturbations in the parameter estimates.

Robust models for the mean and covariance information are a natural way to alleviate this difficulty, and they have been explored by numerous researchers. Lobo and Boyd [98] propose box, ellipsoidal, and other uncertainty sets for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. For example, the box uncertainty sets have the form

$$\begin{aligned} \mathcal{M} &= \left\{ \boldsymbol{\mu} \in \mathbb{R}^n \mid \underline{\mu}_i \leq \mu \leq \overline{\mu}_i, \ i = 1, \ldots, n \right\} \\ \mathcal{S} &= \left\{ \boldsymbol{\Sigma} \in \mathbb{S}^n_+ \mid \underline{\Sigma}_{ij} \leq \Sigma_{ij} \leq \overline{\Sigma}_{ij}, \ i = 1, \ldots, n, \ j = 1, \ldots, n \right\}. \end{aligned}$$

In turn, with these uncertainty structures, they provide a polynomial-time cutting plane algorithm for solving robust variants of Problems (5.15) and (5.16), e.g., the *robust minimum variance problem*

$$\min\left\{\sup_{\boldsymbol{\Sigma}\in\mathcal{S}} \boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w} \;:\; \inf_{\boldsymbol{\mu}\in\mathcal{M}} \boldsymbol{\mu}^\top\boldsymbol{w} \geq r,\; \boldsymbol{w}\in\mathcal{W}\right\}. \tag{5.17}$$

Costa and Paiva [60] propose uncertainty structures of the form $\mathcal{M} = \operatorname{conv}(\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_k)$, $\mathcal{S} = \operatorname{conv}(\boldsymbol{\Sigma}_1,\ldots,\boldsymbol{\Sigma}_k)$, and formulate robust counterparts of (5.15) and (5.16) as optimization problems over linear matrix inequalities.

Tütüncü and Koenig [130] focus on the case of box uncertainty sets for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as well and show that Problem (5.17) is equivalent to the *robust risk-adjusted return problem*

$$\max\left\{\inf_{\boldsymbol{\mu}\in\mathcal{M},\;\boldsymbol{\Sigma}\in\mathcal{S}}\left\{\boldsymbol{\mu}^\top\boldsymbol{w} - \lambda\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}\right\} \;:\; \boldsymbol{w}\in\mathcal{W}\right\}, \tag{5.18}$$

where $\lambda \geq 0$ is an investor-specified risk factor. They are able to show that this is a saddle-point problem, and they use an algorithm of Halldórsson and Tütüncü [86] to compute robust efficient frontiers for this portfolio problem.

5.1.2 Distributional uncertainty models

Less has been said by the Robust Optimization community about *distributional* uncertainty for the return vector in portfolio optimization, perhaps due to the popularity of the classical mean-variance framework of Markowitz. Nonetheless, some work has been done in this regard. Some interesting research on that front is that of El Ghaoui et al. [79], who examine the problem of worst-case *value-at-risk* (VaR) over portfolios with risky returns belonging to a restricted class of probability distributions. The ϵ-VaR for a portfolio $\boldsymbol{w}$ with risky returns $\tilde{\boldsymbol{r}}$ obeying a distribution $\mathbb{P}$ is the optimal value of the problem

$$\min\left\{\gamma \;:\; \mathbb{P}\left(\gamma \leq -\tilde{\boldsymbol{r}}^\top\boldsymbol{w}\right) \leq \epsilon\right\}. \tag{5.19}$$

In turn, the authors in [79] approach the worst-case VaR problem, i.e.,

$$\min\left\{V_{\mathcal{P}}(\boldsymbol{w}) \;:\; \boldsymbol{w}\in\mathcal{W}\right\}, \tag{5.20}$$

where

$$V_{\mathcal{P}}(\boldsymbol{w}) := \left\{\begin{array}{ll}\text{minimize} & \gamma \\ \text{subject to} & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\gamma \leq -\tilde{\boldsymbol{r}}^\top\boldsymbol{w}\right) \leq \epsilon\end{array}\right\}. \tag{5.21}$$

In particular, the authors first focus on the distributional family $\mathcal{P}$ with fixed mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma} \succ \mathbf{0}$. From a tight Chebyshev bound due to Bertsimas and Popescu [32], it was known that (5.20) is equivalent to the SOCP

$$\min \left\{ \gamma \ : \ \kappa(\epsilon)\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\|_2 - \boldsymbol{\mu}^\top \boldsymbol{w} \leq \gamma \right\},$$

where $\kappa(\epsilon) = \sqrt{(1-\epsilon)/\epsilon}$; in [79], however, the authors also show equivalence of (5.20) to an SDP, and this allows them to extend to the case of uncertainty in the moment information. Specifically, when the supremum in (5.20) is taken over all distributions with mean and covariance known only to belong within $\mathcal{U}$, i.e., $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathcal{U}$, [79] shows the following:

1. When $\mathcal{U} = \operatorname{conv}\left((\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \ldots, (\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)\right)$, then (5.20) is SOCP-representable.

2. When $\mathcal{U}$ is a set of component-wise box constraints on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, then (5.20) is SDP-representable.

One interesting extension in [79] is restricting the distributional family to be sufficiently "close" to some reference probability distribution $\mathbb{P}_0$. In particular, the authors show that the inclusion of an entropy constraint

$$\int \log \frac{d\mathbb{P}}{d\mathbb{P}_0} d\mathbb{P} \quad \leq \quad d$$

in (5.20) still leads to an SOCP-representable problem, with $\kappa(\epsilon)$ modified to a new value $\kappa(\epsilon, d)$ (for the details, see [79]). Thus, imposing this smoothness condition on the distributional family only requires modification of the risk factor.

Pinar and Tütüncü [118] study a distribution-free model for near-arbitrage opportunities, which they term *robust profit opportunities*. The idea is as follows: a portfolio $\boldsymbol{w}$ on risky assets with (known) mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is an arbitrage opportunity if (1) $\boldsymbol{\mu}^\top \boldsymbol{w} \geq 0$, (2) $\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} = 0$, and (3) $\boldsymbol{e}^\top \boldsymbol{w} < 0$. The first condition implies an expected positive return, the second implies a guaranteed return (zero variance), and the final condition states that the portfolio can be formed with a negative initial investment (loan).

In an efficient market, pure arbitrage opportunities cannot exist; instead, the authors seek *robust profit opportunities at level* θ, i.e., portfolios $\boldsymbol{w}$ such that

$$\boldsymbol{\mu}^\top \boldsymbol{w} - \theta\sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}} \geq 0, \quad \text{and} \quad \boldsymbol{e}^\top \boldsymbol{x} < 0. \tag{5.22}$$

The rationale for the system (5.22) is similar to the reasoning from Ben-Tal and Nemirovski [15] discussed earlier on approximations to chance constraints. Namely, under some assumptions on the distribution (boundedness and independence across the assets), portfolios that satisfy (5.22) have a positive return with probability at least $1 - e^{-\theta^2/2}$. The authors in [118] then attempt to solve the *maximum-θ robust profit opportunity problem*:

$$\sup_{\theta, \boldsymbol{w}} \left\{ \theta \ : \ \boldsymbol{\mu}^\top \boldsymbol{w} - \theta \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}} \geq 0, \ \boldsymbol{e}^\top \boldsymbol{w} < 0 \right\}. \tag{5.23}$$

They then show that (5.23) is equivalent to a convex quadratic optimization problem and, under mild assumptions, has a closed-form solution.

Along this vein, Popescu [120] has considered the problem of maximizing expected utility in a distributional-robust way when only the mean and covariance of the distribution are known. Specifically, [120] shows that the problem

$$\min_{\boldsymbol{R} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})} \mathbb{E}_{\boldsymbol{R}} \left[u(\boldsymbol{x}' \boldsymbol{R}) \right], \tag{5.24}$$

where u is any utility function and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the mean and covariance, respectively, of the random return $\boldsymbol{R}$, reduces to a three-point problem. [120] then shows how to optimize over this robust objective (5.24) using quadratic programming.

5.1.3 Robust factor models

A common practice in modeling market return dynamics is to use a so-called *factor model* of the form

$$\tilde{\boldsymbol{r}} \quad = \quad \boldsymbol{\mu} + \boldsymbol{V}^\top \boldsymbol{f} + \boldsymbol{\epsilon}, \tag{5.25}$$

where $\tilde{\boldsymbol{r}} \in \mathbb{R}^n$ is the vector of uncertain returns, $\boldsymbol{\mu} \in \mathbb{R}^n$ is an expected return vector, $\boldsymbol{f} \in \mathbb{R}^m$ is a vector of *factor returns* driving the model (these are typically major stock indices or other fundamental economic indicators), $\boldsymbol{V} \in \mathbb{R}^{m \times n}$ is the *factor loading matrix*, and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ is an uncertain vector of residual returns.

Robust versions of (5.25) have been considered by a few authors. Goldfarb and Iyengar [82] consider a model with $\boldsymbol{f} \in \mathcal{N}(\boldsymbol{0}, \boldsymbol{F})$ and $\boldsymbol{\epsilon} \in \mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$, then explicitly account for covariance uncertainty as:

- $\boldsymbol{D} \in \mathcal{S}_d = \left\{ \boldsymbol{D} \mid \boldsymbol{D} = \mathrm{diag}(\boldsymbol{d}), \ d_i \in \left[\underline{d}_i, \overline{d}_i \right] \right\}$
- $\boldsymbol{V} \in \mathcal{S}_v = \{ \boldsymbol{V}_0 + \boldsymbol{W} \mid \|\boldsymbol{W}_i\|_g \leq \rho_i, \ i = 1, \ldots, m \}$
- $\boldsymbol{\mu} \in \mathcal{S}_m = \{ \boldsymbol{\mu}_0 + \boldsymbol{\varepsilon} \mid |\varepsilon|_i \leq \gamma_i, \ i = 1, \ldots, n \}$,

where $\boldsymbol{W}_i = \boldsymbol{W}\boldsymbol{e}_i$ and, for $\boldsymbol{G} \succ \boldsymbol{0}$, $\|\boldsymbol{w}\|_g = \sqrt{\boldsymbol{w}^\top \boldsymbol{G}\boldsymbol{w}}$. The authors then consider various robust problems using this model, including robust versions of the Markowitz problems (5.15) and (5.16), robust Sharpe ratio problems, and robust value-at-risk problems, and show that all of these problems with the uncertainty model above may be formulated as SOCPs. The authors also show how to compute the uncertainty parameters $\boldsymbol{G}$, ρ_i, γ_i, $\underline{d}_i$, $\overline{d}_i$, using historical return data and multivariate regression based on a specific confidence level ω. Additionally, they show that a particular ellipsoidal uncertainty model for the factor covariance matrix $\boldsymbol{F}$ can be included in the robust problems and the resulting problem may still be formulated as an SOCP.

El Ghaoui et al. [79] also consider the problem of robust factor models. Here, the authors show how to compute upper bounds on the robust worst-case VaR problem via SDP for joint uncertainty models in $(\boldsymbol{\mu}, \boldsymbol{V})$ (ellipsoidal and matrix norm-bounded uncertainty models are considered).

5.1.4 Multi-period robust models

The robust portfolio models discussed heretofore have been for single-stage problems, i.e., the investor chooses a *single* portfolio $\boldsymbol{w} \in \mathbb{R}^n$ and has no future decisions. Some efforts have been made on multi-stage problems. Ben-Tal et al. [11] formulate the following, L-stage portfolio problem:

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{i=1}^{n+1} r_i^L x_i^L \\ \text{subject to} & x_i^l = r_i^{l-1} x_i^{l-1} - y_i^l + z_i^l, \ i = 1, \ldots, n, \ l = 1, \ldots, L \\ & x_{n+1}^l = r_{n+1}^{l-1} x_{n+1}^{l-1} + \displaystyle\sum_{i=1}^{n} (1 - \mu_i^l) y_i^l - \sum_{i=1}^{n} (1 + \nu_i^l) z_i^l, \ l = 1, \ldots, L \\ & x_i^l, y_i^l, z_i^l \geq 0, \end{array} \tag{5.26}$$

Here, x_i^l is the dollar amount invested in asset i at time l (asset $n+1$ is cash), r_i^{l-1} is the uncertain return of asset i from period $l-1$ to period l, y_i^l (z_i^l) is the amount of asset i to sell (buy) at the beginning of period l, and μ_i^l (ν_i^l) are the uncertain sell (buy) transaction costs of asset i at period l.

Of course, (5.26) as stated is simply a linear programming problem and contains no reference to the uncertainty in the returns and the transaction costs. The authors note that one can take a multi-stage stochastic programming approach to the problem, but that such an approach may be quite difficult computationally. With tractability in mind, the authors propose an ellipsoidal uncertainty set model (based on the mean of a period's return minus a safety factor θ_l times the standard deviation of that period's return, similar to [118]) for the uncertain parameters, and show how to solve a "rolling horizon" version of the problem via SOCP.

Pinar and Tütüncü [118] explore a two-period model for their robust profit opportunity problem. In particular, they examine the problem

$$\begin{array}{lll} \sup\limits_{\boldsymbol{x}^0} & \inf\limits_{\boldsymbol{r}^1 \in \mathcal{U}} \sup\limits_{\theta, \boldsymbol{x}^1} \theta & \\ \text{subject to} & \boldsymbol{e}^\top \boldsymbol{x}^1 = (\boldsymbol{r}^1)^\top \boldsymbol{x}^0 & \text{(self-financing constraint)} \\ & (\boldsymbol{\mu}^2)^\top \boldsymbol{x}^1 - \theta \sqrt{(\boldsymbol{x}^1)^\top \boldsymbol{\Sigma}_2 \boldsymbol{x}^1} \geq 0 & \\ & \boldsymbol{e}^\top \boldsymbol{x}^0 < 0, & \end{array} \tag{5.27}$$

where $\boldsymbol{x}^i$ is the portfolio from time i to time $i+1$, $\boldsymbol{r}^1$ is the uncertain return vector for period 1, and $(\boldsymbol{\mu}^2, \boldsymbol{\Sigma}_2)$ is the mean and covariance of the return for period 2. The tractability of (5.27) depends critically on $\mathcal{U}$, but [118] derives a solution to the problem when $\mathcal{U}$ is ellipsoidal.

5.1.5 Computational results for robust portfolios

Most of the studies on robust portfolio optimization are corroborated by promising computational experiments. Here we provide a short summary, by no means exhaustive, of some of the relevant results in this vein.

- Ben-Tal et al. [11] provide results on a simulated market model, and show that their robust approach greatly outperforms a stochastic programming approach based on scenarios (the robust has a much lower observed frequency of losses, always a lower standard deviation of returns, and, in most cases, a higher mean return). Their robust approach also compares favorably to a "nominal" approach that uses expected values of the return vectors.

- Goldfarb and Iyengar [82] perform detailed experiments on both simulated and real market data and compare their robust models to "classical" Markowitz portfolios. On the real market data, the robust portfolios did not always outperform the classical approach, but, for high values of the confidence parameter (i.e., larger uncertainty sets), the robust portfolios had superior performance.

- El Ghaoui et al. [79] show that their robust portfolios significantly outperform nominal portfolios in terms of worst-case value-at-risk; their computations are performed on real market data.

- Tütüncü and Koenig [130] compute robust "efficient frontiers" using real-world market data. They find that the robust portfolios offer significant improvement in worst-case return versus nominal portfolios at the expense of a much smaller cost in expected return.

- Erdoğan et al. [69] consider the problems of index tracking and active portfolio management and provide detailed numerical experiments on both. They find that the robust models of Goldfarb and Iyengar [82] can (a) track an index (SP500) with much fewer assets than classical approaches (which has implications from a transaction costs perspective) and (b) perform well versus a benchmark (again, SP500) for active management.

- Delage and Ye [64] consider a series of portfolio optimization experiments with market returns over a six-year horizon. They apply their method, which solves a distribution-robust problem with mean and covariance information based on samples (which they show can be formulated as an SDP) and show that this approach greatly outperforms an approach based on stochastic programming.

- Ben-Tal et al. [6] apply a robust model based on the theory of convex risk measures to a real-world portfolio problem, and show that their approach can yield significant improvements in downside risk protection at little expense in total performance compared to classical methods.

As the above list is by no means exhaustive, we refer the reader to the references therein for more work illustrating the computational efficacy of robust portfolio models.

5.2 Statistics, learning, and estimation

The process of using data to analyze or describe the parameters and behavior of a system is inherently uncertain, so it is no surprise that such problems have been approached from a Robust Optimization perspective. Here we describe some of the prominent, related work.

5.2.1 Robust Optimization and Regularization

Regularization has played an important role in many fields, including functional analysis, numerical computation, linear algebra, statistics, differential equations, to name but a few. Of interest are the properties of solutions to regularized problems. There have been a number of fundamental connections between regularization, and Robust Optimization.

El Ghaoui and Lebret consider the problem of robust least-squares solutions to systems of over-determined linear equations [77]. Given an over-determined system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, an ordinary least-squares problem is $\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|$. In [77], the authors build explicit models to account for uncertainty for the data $[\boldsymbol{A}\ \boldsymbol{b}]$. The authors show that the solution to the ℓ^2-regularized regression problem, is in fact the solution to a robust optimization problem. In particular, the solution to

$$\text{minimize}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|+\rho\sqrt{\|\boldsymbol{x}\|_2^2+1},$$

is also the solution to the robust problem

$$\min_{\boldsymbol{x}} \max_{\|\Delta\boldsymbol{A}\ \Delta\boldsymbol{b}\|_F\leq\rho} \|(\boldsymbol{A}+\Delta\boldsymbol{A})\boldsymbol{x}-(\boldsymbol{b}+\Delta\boldsymbol{b})\|,$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix, i.e., $\|\boldsymbol{A}\|_F=\sqrt{\text{Tr}(\boldsymbol{A}^\top\boldsymbol{A})}$.

This result demonstrates that "robustifying" a solution gives us regularity properties. This has appeared in other contexts as well, for example see [97]. Drawing motivation from the robust control literature, the authors then consider extensions to structured matrix uncertainty sets, looking at the structured robust least-squares (SRLS) problem under linear, and fractional linear uncertainty structure.

In related work, Xu, Caramanis and Mannor [133] consider ℓ^1-regularized regression, commonly called Lasso, and show that this too is the solution to a robust optimization problem. Lasso has been studied extensively in statistics and signal processing (among other fields) due to its remarkable ability to recover sparsity. Recently this has attracted attention under the name of compressed sensing (see [55, 50]). In [133], the authors show that the solution to

$$\text{minimize}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2+\lambda\|\boldsymbol{x}\|_1,$$

is also the solution to the robust problem

$$\min_{\boldsymbol{x}} \max_{\|\Delta\boldsymbol{A}\|_{\infty,2}\leq\rho} \|(\boldsymbol{A}+\Delta\boldsymbol{A})\boldsymbol{x}-\boldsymbol{b}\|,$$

where $\|\cdot\|_{\infty,2}$ is ∞-norm of the 2-norm of the columns. Using this equivalence, they re-prove that Lasso is sparse using a new robust optimization-based explanation of this sparsity phenomenon, thus showing that sparsity is a consequence of robustness.

In [131], the authors consider robust Support Vector Machines (SVM) and show that like Lasso and Tikhonov-regularized regression, norm-regularized SVMs also have a hidden robustness property: their solutions are solutions to a (non-regularized) robust optimization problem. Using this connection, they prove statistical consistency of SVMs without relying on stability or VC-dimension arguments, as past proofs had done. Thus, this equivalence provides a concrete link between good learning properties of an algorithm and its robustness, and provides a new avenue for designing learning algorithms that are consistent and generalize well. For more on this, we refer to the book chapter on Robust Optimization and Machine Learning [54].

5.2.2 Binary classification via linear discriminants

Robust versions of binary classification problems are explored in several papers. The basic problem setup is as follows: one has a collection of data vectors associated with two classes, $\boldsymbol{x}$ and $\boldsymbol{y}$, with elements of both classes belonging to $\mathbb{R}^n$. The realized data for the two classes have empirical means and covariances $(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ and $(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$, respectively. Based on the observed data, we wish to find a linear decision rule for deciding, with high probability, to which class future observations belong. In other words, we wish to find a hyperplane $\mathcal{H}(\boldsymbol{a}, b) = \{\boldsymbol{z} \in \mathbb{R}^n \mid \boldsymbol{a}^\top \boldsymbol{z} = b\}$, with future classifications on new data $\boldsymbol{z}$ depending on the sign of $\boldsymbol{a}^\top \boldsymbol{z} - b$ such that the misclassification probability is as low as possible. (We direct the interested reader to Chapter 12 of Ben-Tal et al. [19] for more discussion on RO in classification problems).

Lanckriet et al. [96] approach this problem first from the approach of distributional robustness. In particular, they assume the means and covariances are known exactly, but nothing else about the distribution is known. In particular, the *Minimax Probability Machine* (MPM) finds a separating hyperplane $(\boldsymbol{a}, b)$ to the problem

$$\max \left\{ \alpha \ : \ \inf_{\boldsymbol{x} \sim (\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)} \mathbb{P}\left(\boldsymbol{a}^\top \boldsymbol{x} \geq b\right) \geq \alpha, \ \inf_{\boldsymbol{y} \sim (\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)} \mathbb{P}\left(\boldsymbol{a}^\top \boldsymbol{y} \leq b\right) \geq \alpha \right\}, \tag{5.28}$$

where the notation $\boldsymbol{x} \sim (\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ means the inf is taken with respect to all distributions with mean $\boldsymbol{\mu}_x$ and covariance $\boldsymbol{\Sigma}_x$. The authors then show that (5.28) can be solved via SOCP. The authors then go on to show that in the case when the means and covariances themselves belong to an uncertainty set defined as follows

$$\mathcal{X} = \left\{ (\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \mid (\boldsymbol{\mu}_x - \boldsymbol{\mu}_x^0)^\top \boldsymbol{\Sigma}_x^{-1} (\boldsymbol{\mu}_x - \boldsymbol{\mu}_x^0) \leq \nu^2, \ \|\boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_x^0\|_F \leq \rho \right\}, \tag{5.29}$$

$$\mathcal{Y} = \left\{ (\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y) \mid (\boldsymbol{\mu}_y - \boldsymbol{\mu}_y^0)^\top \boldsymbol{\Sigma}_y^{-1} (\boldsymbol{\mu}_y - \boldsymbol{\mu}_y^0) \leq \nu^2, \ \|\boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_y^0\|_F \leq \rho \right\}, \tag{5.30}$$

that the problem reduces to an equivalent MPM of the form of (5.28).

Another technique for linear classification is based on so-called *Fisher discriminant analysis* (FDA) [73]. For random variables belonging to class $\boldsymbol{x}$ or class $\boldsymbol{y}$, respectively, and a separating hyperplane $\boldsymbol{a}$, this approach attempts to maximize the Fisher discriminant ratio

$$f(\boldsymbol{a}, \boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_y) := \frac{\left(\boldsymbol{a}^\top (\boldsymbol{\mu}_x - \boldsymbol{\mu}_y)\right)^2}{\boldsymbol{a}^\top \left(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y\right) \boldsymbol{a}}, \tag{5.31}$$

where the means and covariances, as before, are denoted by $(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ and $(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$. The Fisher discriminant ratio can be thought of as a "signal-to-noise" ratio for the classifier, and the discriminant

$$\boldsymbol{a}^{\text{nom}} := \left(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y\right)^{-1} (\boldsymbol{\mu}_x - \boldsymbol{\mu}_y)$$

gives the maximum value of this ratio. Kim et al. [94] consider the *robust Fisher linear discriminant problem*

$$\text{maximize}_{\boldsymbol{a}\neq\boldsymbol{0}} \quad \min_{(\boldsymbol{\mu}_x,\boldsymbol{\mu}_y,\boldsymbol{\Sigma}_x,\boldsymbol{\Sigma}_y)\in\mathcal{U}} f(\boldsymbol{a},\boldsymbol{\mu}_x,\boldsymbol{\mu}_y,\boldsymbol{\Sigma}_x,\boldsymbol{\Sigma}_y), \tag{5.32}$$

where $\mathcal{U}$ is a convex uncertainty set for the mean and covariance parameters. The main result is that if $\mathcal{U}$ is a convex set, then the discriminant

$$\boldsymbol{a}^* \;\; := \;\; \left(\boldsymbol{\Sigma}_x^* + \boldsymbol{\Sigma}_y^*\right)^{-1}(\boldsymbol{\mu}_x^* - \boldsymbol{\mu}_y^*)$$

is optimal to the Robust Fisher linear discriminant problem (5.32), where $(\boldsymbol{\mu}_x^*, \boldsymbol{\mu}_y^*, \boldsymbol{\Sigma}_x^*, \boldsymbol{\Sigma}_y^*)$ is any optimal solution to the convex optimization problem:

$$\min\left\{(\boldsymbol{\mu}_x - \boldsymbol{\mu}_y)^\top(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y)^{-1}(\boldsymbol{\mu}_x - \boldsymbol{\mu}_y) \;:\; (\boldsymbol{\mu}_x,\boldsymbol{\mu}_y,\boldsymbol{\Sigma}_x,\boldsymbol{\Sigma}_y)\in\mathcal{U}\right\}.$$

The result is general in the sense that no structural properties, other than convexity, are imposed on the uncertainty set $\mathcal{U}$.

Other work using robust optimization for classification and learning, includes that of Shivaswamy, Bhattacharyya and Smola [126] where they consider SOCP approaches for handling missing and uncertain data, and also Caramanis and Mannor [53], where robust optimization is used to obtain a model for uncertainty in the label of the training data.

5.2.3 Parameter estimation

Calafiore and El Ghaoui [48] consider the problem of maximum likelihood estimation for linear models when there is uncertainty in the underlying mean and covariance parameters. Specifically, they consider the problem of estimating the mean $\bar{\boldsymbol{x}}$ of an unknown parameter $\boldsymbol{x}$ with prior distribution $\mathcal{N}(\bar{\boldsymbol{x}}, \boldsymbol{P}(\boldsymbol{\Delta}_p))$. In addition, we have an observations vector $\boldsymbol{y} \sim \mathcal{N}(\bar{\boldsymbol{y}}, \boldsymbol{D}(\boldsymbol{\Delta}_d))$, independent of $\boldsymbol{x}$, where the mean satisfies the linear model

$$\bar{\boldsymbol{y}} \;\; = \;\; \boldsymbol{C}(\boldsymbol{\Delta}_c)\bar{\boldsymbol{x}}. \tag{5.33}$$

Given an *a priori* estimate of $\boldsymbol{x}$, denoted by $\boldsymbol{x}_s$, and a realized observation $\boldsymbol{y}_s$, the problem at hand is to determine an estimate for $\bar{\boldsymbol{x}}$ which maximizes the *a posteriori* probability of the event $(\boldsymbol{x}_s, \boldsymbol{y}_s)$. When all of the other data in the problem are known, due to the fact that $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent and normally distributed, the maximum likelihood estimate is given by

$$\bar{\boldsymbol{x}}_{\mathrm{ML}}(\boldsymbol{\Delta}) \;\; = \;\; \arg\min_{\bar{\boldsymbol{x}}} \|F(\boldsymbol{\Delta})\bar{\boldsymbol{x}} - g(\boldsymbol{\Delta})\|^2, \tag{5.34}$$

where

$$
\begin{aligned}
\boldsymbol{\Delta} &= \begin{bmatrix} \boldsymbol{\Delta}_p^\top & \boldsymbol{\Delta}_d^\top & \boldsymbol{\Delta}_c^\top \end{bmatrix}^\top, \\
F(\boldsymbol{\Delta}) &= \begin{bmatrix} \boldsymbol{D}^{-1/2}(\boldsymbol{\Delta}_d)\boldsymbol{C}(\boldsymbol{\Delta}_c) \\ \boldsymbol{P}^{-1/2}(\boldsymbol{\Delta}_p) \end{bmatrix}, \\
g(\boldsymbol{\Delta}) &= \begin{bmatrix} \boldsymbol{D}^{-1/2}(\boldsymbol{\Delta}_d)\boldsymbol{y}_s \\ \boldsymbol{P}^{-1/2}(\boldsymbol{\Delta}_p)\boldsymbol{x}_s \end{bmatrix}.
\end{aligned}
$$

The authors in [48] consider the case with uncertainty in the underlying parameters. In particularly, they parameterize the uncertainty as a linear-fractional (LFT) model and consider the uncertainty set

$$
\boldsymbol{\Delta}_1 = \left\{ \boldsymbol{\Delta} \in \hat{\boldsymbol{\Delta}} \;\middle|\; \|\boldsymbol{\Delta}\| \leq 1 \right\}, \tag{5.35}
$$

where $\hat{\boldsymbol{\Delta}}$ is a linear subspace (e.g., $\mathbb{R}^{p\times q}$) and the norm is the spectral (maximum singular value) norm. The robust or *worst-case maximum likelihood* (WCML) problem, then, is

$$
\text{minimize} \quad \max_{\boldsymbol{\Delta}\in\boldsymbol{\Delta}_1} \|F(\boldsymbol{\Delta})\boldsymbol{x} - g(\boldsymbol{\Delta})\|^2. \tag{5.36}
$$

One of the main results in [48] is that the WCML problem (5.36) may be solved via an SDP formulation. When $\hat{\boldsymbol{\Delta}} = \mathbb{R}^{p\times q}$, (i.e., unstructured uncertainty) this SDP is exact; if the underlying subspace has more structure, however, the SDP finds an upper bound on the worst-case maximum likelihood.

Eldar et al. [68] consider the problem of estimating an unknown, deterministic parameter $\boldsymbol{x}$ based on an observed signal $\boldsymbol{y}$. They assume the parameter and observations are related by a linear model

$$
\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w},
$$

where $\boldsymbol{w}$ is a zero-mean random vector with covariance $\boldsymbol{C}_w$. The *minimum mean-squared error (MSE) problem* is

$$
\min_{\hat{\boldsymbol{x}}} \mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2\right]. \tag{5.37}
$$

Obviously, since $\boldsymbol{x}$ is unknown, this problem cannot be directly solved. Instead, the authors assume some partial knowledge of $\boldsymbol{x}$. Specifically, they assume that the parameter obeys

$$
\|\boldsymbol{x}\|_{\boldsymbol{T}} \leq L, \tag{5.38}
$$

where $\|\boldsymbol{x}\|_{\boldsymbol{T}}^2 = \boldsymbol{x}^\top\boldsymbol{T}\boldsymbol{x}$ for some known, positive definite matrix $\boldsymbol{T} \in \mathbb{S}^n$, and $L \geq 0$. The *worst-case MSE problem* then is

$$
\min_{\hat{\boldsymbol{x}}=\boldsymbol{G}\boldsymbol{y}} \max_{\{\|\boldsymbol{x}\|_{\boldsymbol{T}}\leq L\}} \mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2\right] = \min_{\hat{\boldsymbol{x}}=\boldsymbol{G}\boldsymbol{y}} \max_{\{\|\boldsymbol{x}\|_{\boldsymbol{T}}\leq L\}} \left\{\boldsymbol{x}^\top(\boldsymbol{I}-\boldsymbol{G}\boldsymbol{H})^\top(\boldsymbol{I}-\boldsymbol{G}\boldsymbol{H})\boldsymbol{x} + \mathrm{Tr}(\boldsymbol{G}\boldsymbol{C}_w\boldsymbol{G}^\top)\right\}. \tag{5.39}
$$

Notice that this problem restricts to estimators which are linear in the observations. [68] then shows that (5.39) may be solved via SDP and, moreover, when $\boldsymbol{T}$ and $\boldsymbol{C}_w$ have identical eigenvectors, that the problem admits a closed-form solution. The authors also extend this formulation to include uncertainty in the system matrix $\boldsymbol{H}$. In particular, they show that the robust worst-case MSE problem

$$\min_{\hat{\boldsymbol{x}}=\boldsymbol{G}\boldsymbol{y}} \max_{\{\|\boldsymbol{x}\|_{\boldsymbol{T}}\leq L,\ \|\delta\boldsymbol{H}\|\leq\rho\}} \mathbb{E}\left[\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|^2\right], \tag{5.40}$$

where the matrix $\boldsymbol{H}+\delta\boldsymbol{H}$ is now used in the system model and the matrix norm used is the spectral norm, may also be solved via SDP.

For other work on sparsity and statistics, and sparse covariance estimation, we refer the reader to recent work in [114], [1], and [51].

5.3 Supply chain management

Bertsimas and Thiele [36] consider a robust model for inventory control. They use a cardinality-constrained uncertainty set, as developed in Section 2.2. One main contribution of [36] is to show that the robust problem has an optimal policy which is of the (s_k, S_k) form, i.e., order an amount $S_k - x_k$ if $x_k < s_k$ and order nothing otherwise, and the authors explicitly compute (s_k, S_k). Note that this implies that the robust approach to single-station inventory control has policies which are structurally identical to the stochastic case, with the added advantage that probability distributions need not be assumed in the robust case. A further benefit shown by the authors is that tractability of the problem readily extends to problems with capacities and over networks, and the authors in [36] characterize the optimal policies in these cases as well.

Ben-Tal et al. [9] propose an adaptable robust model, in particular an AARC for an inventory control problem in which the retailer has flexible commitments with the supplier.This model has adaptability explicitly integrated into it, but computed as an *affine* function of the realized demands. Thus, they use the affine adaptable framework of Section 4.3.1 This structure allows the authors in [9] to obtain an approach which is not only robust and adaptable, but also computationally tractable. The model is more general than the above discussion in that it allows the retailer to pre-specify order levels to the supplier (commitments), but then pays a piecewise linear penalty for the deviation of the actual orders from this initial specification. For the sake of brevity, we refer the reader to the paper for details.

Bienstock and Özbay [38] propose a robust model for computing basestock levels in inventory control. One of their uncertainty models, inspired by adversarial queueing theory, is a non-convex model with "peaks" in demand, and they provide a finite algorithm based on Bender's decomposition and show promising computational results.

5.4 Engineering

Robust Optimization techniques have been applied to a wide variety of engineering problems. Many of the relevant references have already been provided in the individual sections above, in particular in Section 2 and subsections therein. In this section, we briefly mention some additional work in this area. For the sake of brevity, we omit most technical details and refer the reader to the relevant papers for more.

Some of the many papers on robust engineering design problems are the following.

1. *Structural design.* Ben-Tal and Nemirovski [12] propose a robust version of a truss topology design problem in which the resulting truss structures have stable performance across a family of loading scenarios. They derive an SDP approach to solving this robust design problem.

2. *Circuit design.* Boyd et al. [41] and Patil et al. [117] consider the problem of minimizing delay in digital circuits when the underlying gate delays are not known exactly. They show how to approach such problems using geometric programming. See also [104, 103, 127], already discussed above.

3. *Power control in wireless channels.* Hsiung et al. [88] utilize a robust geometric programming approach to approximate the problem of minimizing the total power consumption subject to constraints on the outage probability between receivers and transmitters in wireless channels with lognormal fading. For more on applications to communication, particularly the application of geometric programming, we refer the reader to the monograph [58], and the review articles [100, 101]. For applications to coordination schemes and power control in wireless channels, see [135].

4. *Antenna design.* Lorenz and Boyd [99] consider the problem of building an array antenna with minimum variance when the underlying array response is not known exactly. Using an ellipsoidal uncertainty model, they show that this problem is equivalent to an SOCP. Mutapcic et al. [108] consider a beamforming design problem in which the weights cannot be implemented exactly, but instead are known only to lie within a box constraint. They show that the resulting design

problem has the same structure as the underlying, nominal beamforming problem and may, in fact, be interpreted as a regularized version of this nominal problem.

5. *Control.* Notions of robustness have been widely popular in control theory for several decades (see, e.g., Başar and Bernhard [5], and Zhou et al. [136]). Somewhat in contrast to this literature, Bertsimas and Brown [23] explicitly use recent RO techniques to develop a tractable approach to constrained linear-quadratic control problems.

6. *Simulation Based Optimization in Engineering.* In stark contrast to many of the problems we have thus-far described, many engineering design problems do not have characteristics captured by an easily-evaluated and manipulated functional form. Instead, for many problems, the physical properties of a system can often only be described by numerical simulation. In [30], Bertsimas, Nohadani and Teo present a framework for robust optimization in exactly this setting, and describe an application of their robust optimization method for electromagnetic scattering problems.

6 Future directions

The goal of this paper has been to survey the known landscape of the theory and applications of RO. Some of the unknown questions critical to the development of this field are the following:

1. *Tractability of adaptable RO.* While in some very special cases, we have known, tractable approaches to multi-stage RO, these are still quite limited, and it is fair to say that most adaptable RO problems currently remain intractable. The most pressing research directions in this vein, then, relate to tractability, so that a similarly successful theory can be developed as in single-stage static Robust Optimization.

2. *Characterizing the price of robustness.* Some work (e.g., [34, 134]) has explored the cost, in terms of optimality from the nominal solution, associated with robustness. These studies, however, have been largely empirical. Of interest are theoretical bounds to gain an understanding of when robustness is cheap or expensive.

3. *Further developing RO from a data-driven perspective.* While some RO approaches build uncertainty sets directly from data, most of the models in the Robust Optimization literature are not directly connected to data. Recent work on this issue ([64], [24]) have started to lay a foundation

to this perspective. Further developing a data-driven theory of RO is interesting from a theoretical perspective, and also compelling in a practical sense, as many real-world applications are data-rich.

References

[1] F. Bach A. d'Aspremont and L. El Ghaoui. Optimal solutions for sparse principal component analysis. *Journal of Machine Learning Research*, 9:1269–1294, July 2008.

[2] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[3] A. Atamtürk. Strong formulations of robust mixed 0-1 programming. *Mathematical Programming*, 108, 2006.

[4] A. Atamtürk and M. Zhang. Two-stage robust network flow and design under demand uncertainty. *Operations Research*, 55:662–673, 2007.

[5] T. Başar and P. Bernhard. *H^∞-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach.* Birkhäuser, Boston, MA, 1995.

[6] A. Ben-Tal, D. Bertsimas, and D.B. Brown. A soft robust model for optimization under ambiguity. To appear, *Operations Research*, 2009.

[7] A. Ben-Tal, S. Boyd, and A. Nemirovski. Extending the scope of robust optimization. *Math. Programming, Ser. B*, 107:63–89, 2006.

[8] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. Robust semidefinite programming. in Saigal, R., Vandenberghe, L., Wolkowicz, H., eds., *Semidefinite programming and applications*, Kluwer Academic Publishers, 2000.

[9] A. Ben-Tal, B. Golany, A. Nemirovski, and J.P. Vial. Supplier-retailer flexible commitments contracts: A robust optimization approach. *Manufacturing and Service Operations Managment*, 7(3):248–271, 2005.

[10] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Math. Programming*, 99:351–376, 2003.

[11] A. Ben-Tal, T. Margalit, and A. Nemirovski. Robust modeling of multi-stage portfolio problems. *High Performance Optimization*, pages 303–328, 2000.

[12] A. Ben-Tal and A. Nemirovski. Robust truss topology design via semindefinite programming. *SIAM Journal on Optimization*, 7(4):991–1016, 1997.

[13] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Math. Oper. Res.*, 23:769–805, 1998.

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

[15] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Math. Programming*, 88:411–421, 2000.

[16] A. Ben-Tal and A. Nemirovski. On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty. *SIAM Journal on Optimization*, 12:811–833, 2002.

[17] A. Ben-Tal and A. Nemirovski. On approximate robust counterparts of uncertain semidefinite and conic quadratic programs. *Proceedings of 20th IFIP TC7 Conference on System Modelling and Optimization*, July 23-27, 2001, Trier, Germany.

[18] A. Ben-Tal, A. Nemirovski, and C. Roos. Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM Journal on Optimization*, 13(2):535–560, 2002.

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

[20] S. Benati and R. Rizzi. A mixed integer linear programming formulation of the optimal mean/value-at-risk portfolio problem. *European Journal of Operational Research*, 176(1):423–434, 2007.

[21] D. Bertsekas and J.N. Tsitsiklis. *Neuro Dynamic Programming.* Athena Scientific, 1996.

[22] D.P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 1-2. Athena Scientific, Belmont, Mass., 1995.

[23] D. Bertsimas and D.B. Brown. Constrained Stochastic LQC: A Tractable Approach. *IEEE Transactions on Automatic Control*, 52(10):1826–1841, 2007.

[24] D. Bertsimas and D.B. Brown. Constructing uncertainty sets for robust linear optimization. *Operations Research*, 57(6):1483–1495, 2009.

[25] D. Bertsimas and C. Caramanis. Adaptability via sampling. In *Proc. 46th Conference of Decision and Control*, 2007.

[26] D. Bertsimas and C. Caramanis. Finite adaptability in linear optimization. *IEEE Transactions on Automatic Control*, November 2010.

[27] D. Bertsimas and V. Goyal. On the power of robust solutions in two-stage stochastic and adaptive optimization problems. To appear, *Mathematics of Operations Research*, 2009.

[28] D. Bertsimas, D. Iancu, and P. Parrilo. A hierarchy of suboptimal policies for the multi-period, multi-echelon, robust inventory problem. In *INFORMS MSOM*, 2009.

[29] D. Bertsimas, D. Iancu, and P. Parrilo. Optimality of affine policies in multi-stage robust optimization. In *Proc. 48th Conference of Decision and Control*, 2009.

[30] D. Bertsimas, O. Nohadani, and K.M. Teo. Robust optimization in electromagnetic scattering problems. *Journal of Applied Physics*, 101, 2007.

[31] D. Bertsimas, D. Pachamanova, and M. Sim. Robust linear optimization under general norms. *Operations Research Letters*, 32:510–516, 2004.

[32] D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal of Optimization*, 15(3):780–800, 2004.

[33] D. Bertsimas and M. Sim. Robust discrete optimization and network flows. *Mathematical Programming Series B*, 98:49–71, 2003.

[34] D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.

[35] D. Bertsimas and M. Sim. Tractable approximations to robust conic optimization problems. *Mathematical Programming*, 107(1):5–36, 2006.

[36] D. Bertsimas and A. Thiele. A robust optimization approach to supply chain management. *Operations Research*, 54(1):150–168, 2006.

[37] D. Bertsimas and J. Tsitsiklis. *Introduction to Linear Optimization.* Athena Scientific, 1997.

[38] D. Bienstock and N. Özbay. Computing robust basestock levels. *Discrete Optimization*, 5(2):389–414, 2008.

[39] J.R. Birge and F. Louveaux. *Introduction to Stochastic Programming.* Springer-Verlag, 1997.

[40] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory.* Society for Industrial and Applied Mathematics, 1994.

[41] S. Boyd, S.-J. Kim, D. Patil, and M. Horowitz. Digital circuit optimization via geometric programming. *Operations Research*, 53(6):899–932, 2005.

[42] S. Boyd and L. Vandenberghe. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.

[43] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[44] L. Brickman. On the field of values of a matrix. *Proceedings of the AMS*, pages 61–66, 1961.

[45] G. Calafiore and M. C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51:742–753, 2006.

[46] G. Calafiore and M.C. Campi. On two-stage portfolio allocation problems with affine recourse. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 8042–6, 2005.

[47] G. Calafiore and M.C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

[48] G. Calafiore and L. El Ghaoui. Worst-case maximum likelihood estimation in the linear model. *Automatica*, 37(4), April, 2001.

[49] G. Calafiore and B. Polyak. Stochastic algorithms for exact and approximate feasibility of robust lmis. *IEEE Transactions on Automatic Control*, 46(11):1755–1759, 2001.

[50] E. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. Info. Th.*, 51:4203–4215, 2004.

[51] E.J. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? Technical report, Stanford, 2009.

[52] C. Caramanis. PhD dissertation, Massachusetts Institute of Technology, 2006.

[53] C. Caramanis and S. Mannor. Learning in the limit with adversarial disturbances. In *Proceedings of The 21st Annual Conference on Learning Theory*, 2008.

[54] C. Caramanis, S. Mannor, and H. Xu. *Robust Optimization in Machine Learning.* Optimization for Machine Learning. MIT Press, 2011.

[55] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.

[56] X. Chen, M. Sim, and P. Sun. A robust optimization perspective to stochastic programming. *Operations Research*, 55(6):1058–1071, 2007.

[57] X. Chen, M. Sim, P. Sun, and J. Zhang. A linear-decision based approximation approach to stochastic programming. *Operations Research*, 56(2):344–357, 2008.

[58] M. Chiang. Geometric programming for communication systems. *Foundations and Trends in Communications and Information Theory*, 2(1-2):1–154, 2005.

[59] G. Choquet. Theory of capacities. *Ann. Inst. Fourier*, pages 131–295, 1954.

[60] O. Costa and A. Paiva. Robust portfolio selection using linear-matrix inequalities. *J. Econom. Dynam. Control*, 26(6):889–909, 2002.

[61] G.B. Dantzig. Linear programming under uncertainty. *Management Science*, 1(3-4):197–206, 1955.

[62] B. Dean, M.X. Goemans, and J. Vondrák. Approximating the stochastic knapsack problem: The benefit of adaptivity. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science, Rome, Italy*, pages 208–217, 2004.

[63] E. Delage and S. Mannor. Percentile optimization for markov decision processes with parameter uncertainty. To appear, *Operations Research*, 2008.

[64] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. To appear, *Operations Research*, 2009.

[65] C. Dellacherie. Quelque commentaires sur les prolongements de capacitiès. In *Seminaire Probabilities V, Strasbourg, Lecture Notes in Math.* Springer-Verlag.

[66] K. Derinkuyu and M.C. Pinar. On the S-procedure and some variants. *Mathematical Methods of Operations Research*, 64(1):55–77, 2006.

[67] G.E. Dullerud and F. Paganini. *A Course in Robust Control Theory: A Convex Approach*. Springer-Verlag, New York, NY, 1999.

[68] Y. Eldar, A. Ben-Tal, and A. Nemirovski. Robust mean-squared error estimation in the presence of model uncertainties. To appear in *IEEE Trans. on Signal Processing*, 2005.

[69] E. Erdoğan, D. Goldfarb, and G. Iyengar. Robust portfolio management. Technical Report CORC TR-2004-11, IEOR, Columbia University, November 2004.

[70] E. Erdoğan and G. Iyengar. On two-stage convex chance constrained problems. To appear, *Math. Methods of Oper. Res*, 2005.

[71] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1-2):37–61, 2006.

[72] M.K.H. Fan, A. Tits, and J. Doyle. Robustness in the presence of mixed parametric uncertainty and unmodeled dynamics. *IEEE Transactions on Automatic Control*, 36(1):25–38, 1991.

[73] R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

[74] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.

[75] R. Freund. Postoptimal analysis of a linear program under simultaneous changes in matrix coefficients. *Math. Prog. Study*, 24:1–13, 1985.

[76] M. Garey and D. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W.H. Freeman, 1979.

[77] L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM J. Matrix Analysis and Applications*, 18(4):1035–1064, 1997.

[78] L. El Ghaoui and S. Niculescu, editors. *Advances in linear matrix inequality methods in control: advances in design and control*. Society for Industrial and Applied Mathematics, Philadelphia, 2000.

[79] L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.

[80] L. El Ghaoui, F. Oustry, and H. Lebret. Robust solutions to uncertain semidefinite programs. *Siam J. Optimization*, 9(1):33–52, 1998.

[81] M.X. Goemans and J. Vondrák. Stochastic covering and adaptivity. In *Proceedings of LATIN 2006*, pages 532–543, 2006.

[82] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 1:1–38, 2003.

[83] P. J. Goulart and E.C. Kerrigan. Relationships between affine feedback policies for robust control with constraints. In *Proceedings of the 16th IFAC Worls Congress on Automatic Control*, 2005.

[84] P.J. Goulart, E.C. Kerrigan, and J.M. Maciejowski. Optimization over state feedback policies for robust control with constraints. *Automatica*, 42:523–533, 1995.

[85] P. Grieder, P.A. Parrilo, and M. Morari. Robust receding horizon control - analysis and synthesis. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, pages 941–946, 2003.

[86] B.V. Halldórsson and R.H. Tütüncü. An interior-point method for a class of saddle-point problems. *Journal of Optimization Theory and Applications*, 116(3):559–590, 2003.

[87] L. Hansen and T. Sargent. Robust control and model uncertainty. *American Economic Review*, 91:60–66, 2001.

[88] K.-L. Hsiung, S.-J. Kim, and S. Boyd. Power control in lognormal fading wireless channels with uptime probability specifications via robust geometric programming. *Proceedings American Control Conference*, 6:3955–3959, June, 2005.

[89] P. Huber. *Robust Statistics.* Wiley, New York, 1981.

[90] G. Infanger. *Planning under Uncertainty: Solving Large-Scale Stochastic Linear Programs.* Boyd and Fraser, 1994.

[91] G. Iyengar. Robust dynamic programming. *Math. of Operations Research*, 30(2):257–280, 2005.

[92] P. Kall and S. Wallace. *Stochastic Programming.* John Wiley & Sons, 1994.

[93] E.C. Kerrigan and J.M. Maciejowski. On robust optimization and the optimal control of constrained linear systems with bounded state disturbances. In *Proceedings of the European Control Conference*, 2003.

[94] S.-J. Kim, A. Magnani, and S. Boyd. Robust fisher discriminant analysis. *Advances in Neural Information Processing Systems*, 18:659–666.

[95] P. Kouvelis and G. Yu. *Robust discrete optimization and its applications.* Kluwer Academic Publishers, Norwell, MA, 1997.

[96] G.R.G. Lanckriet, L. El Ghaoui, C. Bhattacharyya, and M.I. Jordan. A robust minimax approach to classification. *Journal of Machine Learning Research*, 3:555–582, 2002.

[97] A.S. Lewis. Robust regularization. Technical Report available from: `http://legacy.orie.cornell.edu/∼aslewis/publications/Robust.ps`, Cornell, 2002.

[98] M.S. Lobo and S. Boyd. The worst-case risk of a portfolio. Working paper, 2000.

[99] R. Lorenz and S. Boyd. Robust minimum variance beamforming. *IEEE Transactions on Signal Processing*, 53(5):1684–1696, 2005.

[100] Z.Q. Luo. Applications of convex optimization in signal processing and digital communication. *Mathematical Programming*, 97:177–207, 2003.

[101] Z.Q. Luo and W. Yu. An introduction to convex optimization for communications and signal processing. *IEEE Journal on Selected Areas in Communication*, 24(8):1426–1438, 2006.

[102] F. Maccheroni, M. Marinacci, and A. Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006.

[103] M. Mani, C. Caramanis, and M. Orshansky. Power efficient buffer design using finite adaptable optimization. Technical report, The University of Texas at Austin, 2007.

[104] M. Mani, A. Singh, and M. Orshanksy. Joint design-time and post-silicon minimization on parametric yield loss using adjustable robust optimization. In *Proc. International Conf. on Computer Aided Design.*

[105] H.M. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[106] H.M. Markowitz. *Portfolio Selection.* Wiley, New York, 1959.

[107] S. Mudchanatongsuk, F. Ordó nez, and J. Zhao. Robust solutions for network design under transportation cost and demand uncertainty. Technical report, University of Southern California, 2005.

[108] A. Mutapcic, S.-J. Kim, and S. Boyd. Beamforming with uncertain weights. *IEEE Signal Processing Letters*, 14(5):348–351, 2007.

[109] K. Natarajan, D. Pachamanova, and M. Sim. Constructing risk measures from uncertainty sets. *Operations Research*, 57(5):1129–1141, 2009.

[110] A. Nemirovski. Several NP-hard problems arising in robust stability analysis. *Math. Control Signals Systems*, 6:99–105, 1993.

[111] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17:969–996, 2006.

[112] F. Ordó nez and J. Zhao. Robust capacity expansion of network flows. *Networks*, 50(2).

[113] A. Nilim and L. El Ghaoui. Robust markov decision processes with uncertain transition matrices. *Operations Research*, 53(5), 2005.

[114] O.Banerjee, L. El Ghaoui, and A. d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine Learning Research*, 9:485–516, March 2008.

[115] Y. Oishi and Y. Isaka. Exploiting sparsity in the matrix-dilation approach to robust semidefinite programming. *Journal of the Operations Research Society of Japan*, 52(3):321–338, 2009.

[116] I. Paschalidis and S.-C. Kang. Robust linear optimization: On the benefits of distributional information and applications in inventory control. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 4416–4421, 2005.

[117] D. Patil, S. Yun, S.-J. Kim, A. Cheung, M. Horowitz, and S. Boyd. A new method for design of robust digital circuits. *Proceedings International Symposium on Quality Electronic Design (ISQED)*, pages 676–681, March, 2005.

[118] M. Pinar and R.H. Tütüncü. Robust profit opportunities in risky financial portfolios. *Operations Research Letters*, 33(4):331–340, 2005.

[119] I. Pólik and T. Terlaky. A survey of the s-lemma. *SIAM Review*, 49(3):371–418, 2007.

[120] I. Popescu. Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1):98–112, 2007.

[121] A. Prékopa. *Stochastic Programming.* Kluwer, 1995.

[122] M. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley-Interscience, 1994.

[123] J. Renegar. Some perturbation theory for linear programming. *Mathematical Programming*, 65:73–91, 1994.

[124] D. Schmeidler. Integral representation without additivity. In *Proceedings of the American Math Society*, volume 97(2), pages 255–261, 1986.

[125] A. Shapiro. On complexity of multistage stochastic programs. *Operations Research Letters*, 34:1–8, 2006.

[126] P. Shivaswamy, C. Bhattacharyya, and A.J. Smola. Second order cone programming approaches for handling missing and uncertain data. *Journal of Machine Learning Research*, 7:1283–1314, 2006.

[127] A. Singh, K. He, C. Caramanis, and M. Orshansky. Mitigation of intra-array sram variability using adaptive voltage architecture. In *Proc. International Conference on Computer Aided Design (ICCAD)*, 2009.

[128] A.L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21:1154–1157, 1973.

[129] A. Takeda, S. Taguchi, and R. Tütüncü. Adjustable robust optimization models for nonlinear multi-period optimization. To appear, *Journal of Optimization Theory and Applications*, 2004.

[130] R.H. Tütüncü and M. Koenig. Robust asset allocation. *Annals of Operations Research*, 132:157–187, 2004.

[131] H. Xu, C. Caramanis, and S. Mannor. Robustness and regularization of support vector machines. *Journal of Machine Learning Research*, 10:1485–1510, 2009.

[132] H. Xu, C. Caramanis, and S. Mannor. A distributional interpretation of robust optimization. In *Proceedings of the Allerton Conference on Communication, Control and Computing*, 2010.

[133] H. Xu, C. Caramanis, and S. Mannor. Robust regression and Lasso. *IEEE Transactions on Information Theory*, 56(7):3561–3574, 2010.

[134] H. Xu and S. Mannor. Tradeoff performance and robustness in linear programming and markov decision processes. In *NIPS*, 2006.

[135] S. Yun and C. Caramanis. System level optimization in wireless networks with uncertain customer arrival rates. In *Proc. Allerton Conference on Communications, Control and Computing*, 2008.

[136] K. Zhou, J.C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice-Hall, Upper Saddle River, NJ, 1996.