Theory and applications of Robust Optimization

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

July 6, 2007

Abstract

In this paper we survey the primary research, both theoretical and applied, in the field of Robust Optimization (RO). Our focus will be on the computational attractiveness of RO approaches, as well as the modeling power and broad applicability of the methodology. In addition to surveying the most prominent theoretical results of RO over the past decade, we will also present some recent results linking RO to adaptable models for multi-stage decision-making problems. Finally, we will highlight successful applications of RO across a wide spectrum of domains, including, but not limited to, finance, statistics, learning, and engineering.

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

1 Introduction

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

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 [44]. We refer the interested reader to several textbooks ([64, 31, 87, 66]) 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.

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

†Assistant Professor of Decision Sciences, 1 Towerview Drive, Fuqua School of Business, Duke University, Durham, NC 27705. dbbrown@duke.edu

‡Assistant Professor, Department of Electrical and Computer Engineering, The University of Texas at Austin, 1 University Station, Austin, TX 78712. cmcaram@ece.utexas.edu

Instead of seeking to immunize the solution in some probabilistic sense to stochastic uncertainty, here the decision-maker constructs a solution that is optimal 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 the most appropriate notion of parameter uncertainty. Next, 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.

In the early 1970s, Soyster [92] was one of the first researchers to investigate explicit approaches to Robust Optimization. This short note focused on robust linear optimization in the case where the column vectors of the constraint matrix were constrained to belong to ellipsoidal uncertainty sets; Falk [50] followed this a few years later with more work on "inexact linear programs." The optimization community, however, was relatively quiet on the issue of robustness until the work of Ben-Tal and Nemirovski (e.g., [13, 14, 15]) and El Ghaoui et al. [56, 58] in the late 1990s. This work, 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, [34]) sparked a massive flurry of interest in the field of Robust Optimization.

Central issues we seek to address in this paper include:

1. Tractability of Robust Optimization models: In particular, given a class of nominal problems (e.g., LP, SOCP, SDP, etc.) and a structured uncertainty set (polyhedral, ellipsoidal, etc.), what is the complexity class of the corresponding robust problem?

2. Conservativeness and probability guarantees: How much flexibility does the designer have in selecting the uncertainty sets? What guidance does he have for this selection? And what do these uncertainty sets tell us about probabilistic feasibility guarantees under various distributions for the uncertain parameters?

3. Flexibility, applicability, and modeling power: What uncertainty sets are appropriate for a given application? How fragile are the tractability results? For what applications is this general methodology suitable?

As a preview of what is to come, we give (abdridged) answers to the three issues raised above.

1. Tractability: In general, the robust version of a tractable optimization problem may not itself be tractable. In this paper 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.

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 no distributional information is available. But even if there is an underlying distribution, the tractability benefits of the Robust Optimization paradigm 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.

3. Flexibility and modeling power: In Section 2, we survey 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 broad variety of applications.

The overall aim of this paper is to outline the development and main aspects of Robust Optimization, with an emphasis on its power, flexibility, and structure. We will also highlight some exciting and important open directions of current research, as well as the broad applicability of RO. Section 2 focuses on the structure and tractability of the main results, describing when, where, and how Robust Optimization is applicable. Section 3 describes important new directions in Robust Optimization, in particular multi-stage adaptable Robust Optimization, which is much less developed, and rich with open questions. In Section 4, we detail a wide spectrum of application areas to illustrate the broad impact that Robust Optimization has had in the early part of its development.

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, and we present some *a priori* probabilistic feasibility and optimality guarantees for solutions to Robust Optimization problems.

2.1 Robust Optimization

The general Robust Optimization formulation is:

$$\begin{array}{lll} \text{minimize} & f_0(\boldsymbol{x}) & \\ \text{subject to} & f_i(\boldsymbol{x}, \boldsymbol{u}_i) \leq 0, & \forall\, \boldsymbol{u}_i \in \mathcal{U}_i,\ i = 1, \ldots, m. \end{array} \tag{2.1}$$

Here $\boldsymbol{x} \in \mathbb{R}^n$ is a vector of decision variables, f_0, f_i are as before, $\boldsymbol{u}_i \in \mathbb{R}^k$ are *disturbance vectors* or *parameter uncertainties*, and $\mathcal{U}_i \subseteq \mathbb{R}^k$ are *uncertainty sets*, which, for our purposes, will always be closed.

Note that by introducing a new constraint if necessary, we can always take the objective function to have no uncertainty. The goal of (2.1) 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 $\mathcal{U}_i$. If $\mathcal{U}_i$ is a singleton, then the corresponding constraint has no uncertainty. 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 (2.1):

- We can assume without loss of generality that the uncertainty set $\mathcal{U}$ has the form $\mathcal{U} = \mathcal{U}_1 \times \ldots \times \mathcal{U}_m$, due to the constraint-wise feasibility requirements (see also Ben-Tal and Nemirovski, [14]).

- Problem (2.1) also contains the instances when the decision or disturbance vectors are contained in more general vector spaces than $\mathbb{R}^n$ or $\mathbb{R}^k$, such as $\mathbb{S}^n$ in the case of semidefinite optimization.

We emphasize that Robust Optimization is distinctly different than *sensitivity analysis*, which is typically applied as a post-optimization tool for quantifying the change in cost for small perturbations in the underlying problem data. Here, our goal is to *compute* solutions with *a priori* ensured feasibility when the problem parameters vary within the prescribed uncertainty set. We refer the reader to some of the standard optimization literature (e.g., Bertsimas and Tsitsiklis, [29], Boyd and Vandenberghe, [35]) and works on perturbation theory (e.g., Freund, [53], Renegar, [88]) for more on sensitivity analysis.

It is not at all clear when (2.1) is efficiently solvable, since as written (2.1) may have infinitely many constraints. In general, the robust problem is intractable, however, manyu interesting classes of problems admit efficient solution. and much of the literature since the modern resurgence has focused on specifying classes of functions f_i, coupled with the types of uncertainty sets $\mathcal{U}_i$, that yield tractable problems. If we define the robust feasible set to be

$$X(\mathcal{U}) \quad = \quad \{\boldsymbol{x} \mid f_i(\boldsymbol{x}, \boldsymbol{u}_i) \leq 0 \ \forall \ \boldsymbol{u}_i \in \mathcal{U}_i, \ i = 1, \ldots, m\}, \tag{2.2}$$

then for the most part, tractability is tantamount to $X(\mathcal{U})$ being convex in $\boldsymbol{x}$, with an efficiently computable membership test. We now present an abridged taxonomy of some of the main results.

2.2 An Example: Robust Inventory Control

In order to motivate subsequent developments, we give an example to inventory control with demand uncertainty (see Adida and Perakis [1], Bertsimas and Thiele [28], Ben-Tal et al. [10], and references therein) in order to motivate developments in the sequel. We revisit this example in more detail in Section 4. The essence of the problem is to make ordering, stocking, and storage decisions in order to meet demand, so that the cost is minimized. Cost is incurred from the actual purchases including fixed costs of placing an order, but also from holding and shortage costs. The basic stock evolution equation is given by: $x_{k+1} = x_k + u_k - w_k, \quad k = 0, \ldots, T-1$, where u_k is the stock ordered at the beginning of the k^{th} period, and w_k is the demand during that same period. Assuming that we incur a holding cost

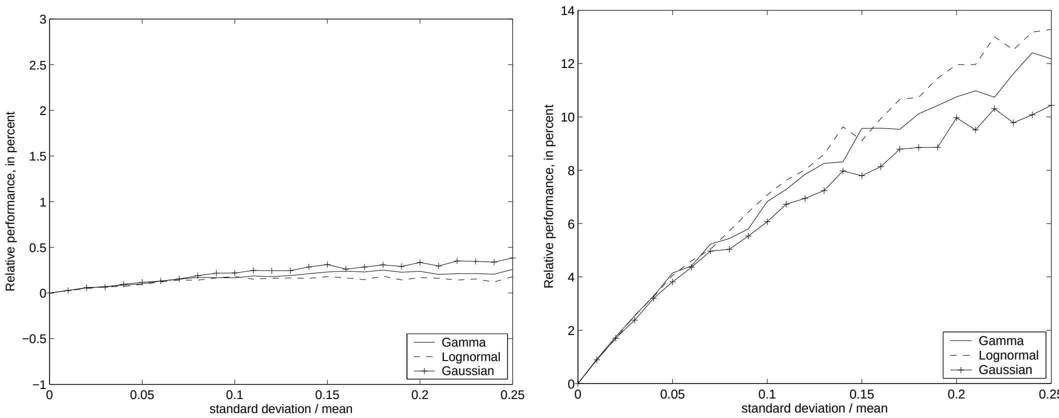

Figure 1: These figures show the relative performance of dynamic and robust optimization for three distributions of the demand: Gamma, Lognormal, and Gaussian. The figure on the left shows the case where the distribution of the demand uncertainty is known exactly; the figure on the right assumes that only the first two moments are known exactly.

(extra stock) hx, and shortage cost $-px$, this can be written as $y = \max\{hx, -px\}$. When the demands are known deterministically, we can write the optimal T-stage inventory control problem as:

$$\begin{aligned} \min : \quad & \sum_{k=0}^{T-1} (cu_k + Kv_k + y_k) \\ \text{s.t. :} \quad & y_k \geq h\left(x_0 + \sum_{i=0}^{k}(u_i - w_i)\right), \quad k = 0, \dots, T-1, \\ & y_k \geq -p\left(x_0 + \sum_{i=0}^{k}(u_i - w_i)\right), \quad k = 0, \dots, T-1, \\ & 0 \leq u_k \leq Mv_k, \ v_k \in \{0,1\}, \quad k = 0, \dots, T-1. \end{aligned}$$

Here, v_k denotes the decision to purchase or not during period k, and is only required if there is a fixed cost for ordering. M is the upper bound on the order size.

Dynamic programming approaches for dealing with uncertainty of w_k assume knowledge of the distribution of the w_k; furthermore, their tractability depends on the particular distribution, and special structure of the problem. In particular, extending them from the single-station case presented here, to the network case, appears to be intractable. The ideas presented in this paper propose modeling the demand-uncertainty deterministically, choosing uncertainty sets rather than distributions. The graphs in Figure **??** show the simulated relative perfomance of the dynamic programming solution to the robust optimization solution, when the assumed and actual distributions of the demands are identical, and then under the much more realistic assumption that they are known only up to their first two moments. In the former case, the performance is essentially identical; in the latter case, we see that as the standard deviation increases, the robust optimization policy outperforms dynamic programming by 10-13%. For full details on the simulations, see [28].

Immediate questions include: What is the complexity, and structure of the resulting robust problem for different classes of uncertainty set $\mathcal{U}$? Fixed costs result in a mixed integer optimization problem. When can robust optimization techniques address this class of problems? How can we control conservativeness via a "budget of uncertainty"?

2.3 Robust linear optimization

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

$$\begin{array}{ll} \text{minimize} & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}, \qquad \forall\, \boldsymbol{a_1} \in \mathcal{U}_1, \ldots, \boldsymbol{a_m} \in \mathcal{U}_m, \end{array} \tag{2.3}$$

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, $\boldsymbol{a}_i^\top \boldsymbol{x} \le b_i$, $\forall \boldsymbol{a}_i \in \mathcal{U}_i$, if and only if

$$\max_{\{\boldsymbol{a}_i \in \mathcal{U}_i\}} \boldsymbol{a}_i^\top \boldsymbol{x} \le b_i, \quad \forall\, i. \tag{2.4}$$

We refer to this as the *subproblem* which must be solved; its structure determines the complexity of solving the Robust Optimization problem.

Ellipsoidal Uncertainty: Ben-Tal and Nemirovski [14], as well as El Ghaoui et al. [56, 58], consider ellipsoidal uncertainty sets, in part motivated by the normal distribution. 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.

Theorem 1. *([14]) Let $\mathcal{U}$ be "ellipsoidal," i.e., $\mathcal{U} = U(\Pi, \boldsymbol{Q}) = \{\Pi(\boldsymbol{u}) \mid \|\boldsymbol{Q}\boldsymbol{u}\| \le \rho\}$, 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 (2.3) is equivalent to a second-order cone program (SOCP). Explicitly, if we have the uncertain optimization*

$$\begin{array}{ll} \textit{minimize} & \boldsymbol{c}^\top \boldsymbol{x} \\ \textit{subject to} & \boldsymbol{a}_i \boldsymbol{x} \le 0, \quad \forall \boldsymbol{a}_i \in \mathcal{U}_i, \quad \forall i = 1, \ldots, m, \end{array}$$

where the uncertainty set is given as:

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

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

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

The intuition is as follows: for ellipsoidal uncertainty, the subproblem (2.4) 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]. In fact, 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:

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

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:

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

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 ([26]) use this duality with a family of polyhedral sets that encode a budget of uncertainty in terms of cardinality constraints, as opposed to size of an ellipsoid. The uncertainty sets they consider control the number of parameters of the problem that are allowed to vary from their nominal values, providing a different trade-off between the optimality of the solution, and its robustness to parameter perturbation. In [23], the authors show that these cardinality constrained uncertainty sets can be expressed as norm-bounded uncertainty sets.

The cardinality constrained uncertainty sets are as follows. Given an uncertain matrix, $\boldsymbol{A} = (a_{ij})$, suppose that in row i, the entries a_{ij} for $j \in J_i \subseteq \{1, \ldots, n\}$ vary in some interval about their nominal value, $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$. Rather than protect against the case when every parameter can deviate, as in the original model of Soyster ([92]), we allow at most Γ_i coefficients to deviate. Thus the positive number Γ_i denotes the budget of uncertainty for the i^{th} constraint.[1] Given values $\Gamma_1, \ldots, \Gamma_m$, the robust formulation becomes:

$$\begin{aligned} \min: \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.}: \quad & \textstyle\sum_j a_{ij} x_j + \max_{\{S_i \subseteq J_i \,:\, |S_i| = \Gamma_i\}} \sum_{j \in S_i} \hat{a}_{ij} y_j \leq b_i \quad & 1 \leq i \leq m \\ & -y_j \leq x_j \leq y_j & 1 \leq j \leq n \\ & \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u}, \ \ \boldsymbol{y} \geq \boldsymbol{0}. \end{aligned} \tag{2.5}$$

[1]For the full details see [26].

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):

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

Norm Uncertainty: Bertsimas et al. [23] show that robust linear optimization problems with uncertainty sets described by more general norms lead to convex problems with constraints related to the dual norm. We use $\text{vec}(\boldsymbol{A})$ to denote the vector formed by concatenating the rows of the matrix $\boldsymbol{A}$.

Theorem 2. *(Bertsimas et al., [23]) 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 (2.3) 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.

2.4 Robust quadratic optimization

For $f_i(\boldsymbol{x}, \boldsymbol{u}_i)$ of the form

$$\|\boldsymbol{A}_i\boldsymbol{x}\|^2 + \boldsymbol{b}_i^\top \boldsymbol{x} + c_i \le 0,$$

i.e., (convex) *quadratically constrained quadratic programs* (QCQP), where $\boldsymbol{u}_i = (\boldsymbol{A}_i, \boldsymbol{b}_i, c_i)$, the robust counterpart is a semidefinite optimization problem if $\mathcal{U}$ is a single ellipsoid, and NP-hard if $\mathcal{U}$ is polyhedral (Ben-Tal and Nemirovski, [13, 14]).

For robust SOCPs, the $f_i(\boldsymbol{x}, \boldsymbol{u}_i)$ are of the form

$$\|\boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{b}_i\| \le \boldsymbol{c}_i^\top \boldsymbol{x} + d_i.$$

If $(\boldsymbol{A}_i, \boldsymbol{b}_i)$ and $(\boldsymbol{c}_i, d_i)$ each belong to a set described by a single ellipsoid, then the robust counterpart is a semidefinite optimization problem; if $(\boldsymbol{A}_i, \boldsymbol{b}_i, \boldsymbol{c}_i, d_i)$ varies within a shared ellipsoidal set, however, the robust problem is NP-hard (Ben-Tal et al., [18], Bertsimas and Sim, [27]).

We illustrate here only how to obtain the explicit reformulation of a robust quadratic constraint, subject to simple ellipsoidal uncertainty.[2] We follow Ben-Tal, Nemirovski and Roos ([18]). Consider the quadratic constraint

$$\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} \leq 2\boldsymbol{b}^\top \boldsymbol{x} + \boldsymbol{c}, \quad \forall (\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}) \in \mathcal{U}, \tag{2.6}$$

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

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

A vector $\boldsymbol{x}$ is feasible for the robust constraint (2.6) if and only if it is feasible for the constraint:

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

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. While this problem is not convex, it can be reformulated as a (convex) semidefinite optimization problem.[3]

If the uncertainty set is an intersection of ellipsoids, then exact solution of the subproblem is NP-hard.[4] We return to this in Section 3 where we consider multistage optimization.

For the single ellipsoid case, our original problem of feasibility for the robustified quadratic constraint becomes an SDP feasibility problem. Therefore subject to mild regularity conditions (e.g., Slater's condition) strong duality holds, and by using the dual to the SDP, 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 (2.6) if and only if there exists a scalar $\tau \in \mathbb{R}$ such that the following matrix inequality holds:*

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

2.5 Robust Semidefinite Optimization

With ellipsoidal uncertainty sets, robust counterparts of semidefinite optimization problems are 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, [79]). Computing approximate solutions that are robust *feasible*

[2] Here, *simple ellipsoidal uncertainty* means the uncertainty set is a single ellipsoid, as opposed to an intersection of several ellipsoids.

[3] Related to this and also well-known, is the so-called S-lemma (or S-procedure) in control (e.g., Boyd et al. [32]).

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

but not robust *optimal* to robust semidefinite optimization problems has, as a consequence, received considerable attention (e.g., [58], [17, 16], and [27]). 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 to the feasible set is to the true feasible set. Precisely, the measure for this is: $\rho(\mathrm{AR} : \mathrm{R}) = \inf\{\rho \geq 1 \mid X(\mathrm{AR}) \supseteq X(\mathcal{U}(\rho))\}$, 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. Ben-Tal and Nemirovski develop an inner approximation ([17]) such that $\rho(\mathrm{AR} : \mathrm{R}) \leq \pi\sqrt{\mu}/2$, where μ is the maximum rank of the matrices describing $\mathcal{U}$.

2.6 Robust geometric programming

A *geometric program* (GP) is a convex optimization problem of the form

$$\begin{array}{ll} \text{minimize} & \boldsymbol{c}^\top \boldsymbol{y} \\ \text{subject to} & g(\boldsymbol{A}_i \boldsymbol{y} + \boldsymbol{b}_i) \leq 0, \qquad i = 1, \ldots, m, \\ & \boldsymbol{G}\boldsymbol{y} + \boldsymbol{h} = \boldsymbol{0}, \end{array}$$

where $g : \mathbb{R}^k \to \mathbb{R}$ is the *log-sum-exp* function, $g(\boldsymbol{x}) = \log\left(\sum_{i=1}^{k} e^{x_i}\right)$, and the matrices and vectors $\boldsymbol{A}_i$, $\boldsymbol{G}$, $\boldsymbol{b}_i$, and $\boldsymbol{h}$ are of appropriate dimension. For many engineering, design, and statistical applications of GP, see Boyd and Vandenberghe [35]. Hsiung et al. [61] study a robust version of GP with constraints

$$g(\tilde{\boldsymbol{A}}_i(\boldsymbol{u})\boldsymbol{v} + \tilde{\boldsymbol{b}}_i(\boldsymbol{u})) \quad \leq \quad 0 \qquad \forall\, \boldsymbol{u} \in \mathcal{U},$$

where $(\tilde{\boldsymbol{A}}_i(\boldsymbol{u}), \tilde{\boldsymbol{b}}_i(\boldsymbol{u}))$ are affinely dependent on the uncertainty $\boldsymbol{u}$, and $\mathcal{U}$ is an ellipsoid or a polyhedron. The complexity of this problem is unknown; the approach in [61] is to use a piecewise linear approximation to get upper and lower bounds to the robust GP.

2.7 Robust discrete optimization

Kouvelis and Yu [68] 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 [68].

Bertsimas and Sim [25], 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. They show that 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. They also show that this result extends to approximation algorithms for combinatorial problems. For network flow problems, they show that 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 mixed 0-1 problem with size polynomial in the size of a tight linear programming formulation for the nominal mixed 0-1 problem.

2.8 Robust convex optimization

The robust counterpart to a general conic convex optimization problem is typically nonconvex and intractable ([13]). This is implied by the results described above, since conic problems include semidefinite optimization. Nevertheless, there are some approximate formulations of the general conic convex robust problem. We refer the interested reader to the recent work by Bertsimas and Sim [27].

2.9 Probability guarantees

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 $\boldsymbol{x} \in X(\mathcal{U})$ implies $\mathbb{P}\left(f_i(\boldsymbol{x}, \boldsymbol{u}_i) > 0\right) \leq \epsilon$, under (ideally mild) assumptions on a distribution for $\boldsymbol{u}_i$? In this section, we briefly survey some of the results in this vein.

For linear optimization, Ben-Tal and Nemirovski [15] propose a robust model based on ellipsoids of radius Ω. Under this model, if the uncertain coefficients have bounded, symmetric support, they show that the corresponding robust feasible solutions are feasible with probability $e^{-\Omega^2/2}$. In a similar spirit, Bertsimas and Sim [26] propose an uncertainty set of the form

$$\mathcal{U}_\Gamma = \left\{ \bar{\boldsymbol{A}} + \sum_{j \in J} z_j \hat{a}_j \;\middle|\; \|\boldsymbol{z}\|_\infty \leq 1,\ \sum_{j \in J} \mathbf{1}(z_j) \leq \Gamma \right\}, \tag{2.7}$$

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 $\Gamma \leq |J|$ 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 above. The following then holds.

Theorem 4. *(Bertsimas and Sim [26]) Let $\boldsymbol{x}^*$ satisfy the constraint*

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

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

where $\mathcal{U}_\Gamma$ is as in (2.7). 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

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

In the case of linear optimization with only partial moment information (specifically, known mean and covariance), Bertsimas et al. [23] 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 [23] to imply the guarantee

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

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

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 [27]. See also the work of Chen, Sim, and Sun, in [41], where more general deviation measures are considered, leading to improved probability guarantees. Paschalidis and Kang on probability guarantees and uncertainty set selection when the *entire* distribution is available [84].

2.10 Constructing uncertainty sets

In terms of how to construct uncertainty sets, much of the RO literature assumes an underlying structure *a priori*, then chooses from a parameterized family based on some notion of conservatism (e.g., probability guarantees in the previous section). This is proposed, e.g., in [23, 26, 27]. For instance, one could use a norm-based uncertainty model as explained above. All that is left is to choose the parameter Ω, and this may be done to meet a probability guarantee suitable for the purposes of the decision-maker.

Such an approach assumes a fixed, underlying structure for the uncertainty set. In contrast to this, Bertsimas and Brown [20] connect uncertainty sets to *risk preferences* for the case of linear optimization. In particular, they show that when the decision-maker can express risk preferences for satisfying feasibility in terms of a *coherent risk measure* (Artzner et al., [2]), then an uncertainty set with an explicit construction naturally arises. A converse result naturally holds as well; that is, every uncertainty set coincides with a particular coherent risk measure (Natarajan et al. [78] consider this problem of risk preferences implied by uncertainty sets in detail). Thus, for the case of robust linear optimization, uncertainty sets and risk measures have a one-to-one correspondence.

Ben-Tal, Bertsimas and Brown [6] extend this correspondence to more general risk measures called *convex risk measures* (see, e.g., Föllmer and Schied, [52]) and find a more flexible notion of robustness, where one allows varying degrees of feasibility for different realizations of the uncertain parameters.

3 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 the uncertainty is realized. In many problems, this single-shot assumption may be too restrictive and conservative. We consider here ways to remove it.

Consider the inventory control example from Section **??**, with a single product, one warehouse, and I factories (see [10]). Let $d(t)$ be the demand for the product at time t, only approximately known: $d(t) \in [d_t^* - \theta d_t^*, d_t^* + \theta d_t^*]$. Varying θ, we can model different prediction accuracies for the demand. Let $v(t)$ be the amount of the product in the warehouse at time t. The decision variables are $u(i,t)$, the amount ordered at period t from factory i, and the cost is $c(i,t)$. Finally, let $U(i,t)$ be the production cap on factory i at period t, and $U_T(i)$ the total production cap on factory i. In this adaptable setting, the ordering decisions are made over time, and thus depend on some subset of the past realizations of the demand. Let $D(t)$ denote the set of demand realizations available when the period t ordering decisions are made (so if $D(t) = \emptyset$, then we recover the static setup). Then, the inventory control problem becomes:

$$
\begin{array}{ll}
\min: & F \\
\text{s.t.}: & \displaystyle\sum_{t=1}^{T}\sum_{i=1}^{I} c_i(t) p_i(t, D(t)) \leq F \\
& 0 \leq p_i(t, D(t)) \leq P_i(t), \quad i = 1, \ldots, I, \; t = 1, \ldots, T \\
& \displaystyle\sum_{t=1}^{T} p_i(t, D(t)) \leq Q(i), \quad i = 1, \ldots, I \\
& v(t+1) = v(t) + \displaystyle\sum_{i=1}^{I} p_i(t, D(t)) - d_t, \quad t = 1, \ldots, T \\
& \forall d(t) \in [d_t^* - \theta d_t^*, d_t^* + \theta d_t^*], \quad t = 1, \ldots, T.
\end{array}
$$

We discuss several ways to model the dependency of $p_i(t, D(t))$ on $D(t)$. In particular, [10] considers affine dependence on $D(t)$, and they show that in this case, the inventory problem above can be reformulated as a linear optimization. In particular, they compare their affine approach to two extremes: the static problem, where all decisions are made at the initial time, and the utopic (perfect foresight) solution, where the demand realization is assumed to be known non-causally. For a 24-period example with 3 factories, and sinusoidally varying demand (to model seasonal variations) $d_t^* = 1000\left(1 + \frac{1}{2}\sin\left(\frac{\pi(t-1)}{12}\right)\right)$, they find that the dynamic formulation with affine functions, is comparable to the utopic solution, greatly improving upon the static solution. We report these results in Table 1 (for the full details, see [10]).

Inventory control problems are just one example of multi-stage optimization. Portfolio management problems with multiple investment rounds are another example ([11], and see more on this in Section 4). Other application examples include network design ([4, 80]), dynamic scheduling problems in air traffic control ([39, 81, 83]) and traffic scheduling, and also problems from engineering, such as integrated circuit design with two fabrication stages ([73, 72]).

In this section, we discuss several RO-based approaches to the multi-stage setting.

	2.5% Uncertainty	5% Uncertainty	10% Uncertainty
Static:	4.3%	infeasible	infeasible
Affine:	0.3%	0.6%	1.6%

Table 1: Multi-period inventory control: static and affine adaptable vs the utopic solution.

3.1 Motivation and Background

This section focuses primarily on the linear case. Consider a generic 3-stage linear problem:

$$\begin{array}{rl} \min: & \boldsymbol{c}^\top \boldsymbol{x}_1 \\ \text{s.t.}: & \boldsymbol{A}_1(\boldsymbol{u}_1,\boldsymbol{u}_2)\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{u}_1,\boldsymbol{u}_2)\boldsymbol{x}_2(\boldsymbol{u}_1) + \boldsymbol{A}_3(\boldsymbol{u}_1,\boldsymbol{u}_2)\boldsymbol{x}_3(\boldsymbol{u}_1,\boldsymbol{u}_2) \leq \boldsymbol{b}, \quad \forall(\boldsymbol{u}_1,\boldsymbol{u}_2) \in \mathcal{U}. \end{array} \tag{3.8}$$

Note that we can assume only $\boldsymbol{x}_1$ appears in the cost function, without loss of generality. The sequence of events, reflected in the functional dependencies written in, is as follows: 1a. Decision $\boldsymbol{x}_1$ is implemented. 1b. Uncertainty parameter $\boldsymbol{u}_1$ is realized. 2a. Decision $\boldsymbol{x}_2$ is implemented, after $\boldsymbol{x}_1$ has been implemented, and $\boldsymbol{u}_1$ realized and observed. 2b. Uncertainty parameter $\boldsymbol{u}_2$ is realized. 3. The final decision $\boldsymbol{x}_3$ is implemented, after $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ have been implemented, and $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ realized and observed.

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.

3.1.1 Folding Horizon, Stochastic Optimization, and Dynamic Programming

The most straightforward extension of the single-shot Robust Optimization formulation to that of sequential decision-making, is the so-called folding horizon approach, akin to open-loop feedback in control. Here, the static solution over all stages is computed, and the first-stage decision is implemented. At the next stage, the process is repeated. This algorithm may be quite conservative, as it does not explicitly build into the computation the fact that at the next stage the computation will be repeated with potentially additional information about the uncertainty.

In Stochastic Optimization, the multi-stage formulation has long been a topic of research, particularly for the case of complete recourse. There are approaches using chance constraints, as well as using violation penalties, and we refer the reader to references cited previously for more on this.

Sequential decision-making under uncertainty has traditionally been the domain of Dynamic Programming ([19]). This has recently been extended to the robust Dynamic Programming and robust MDP setting, where the payoffs and the dynamics are not exactly known, in Iyengar [65] and Nilim and El Ghaoui [82], and then also in Huan and Mannor [63]. 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 [65, 82, 63, 45] 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.

3.2 Theoretical Results

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. Nevertheless, there has recently been considerable effort devoted to obtaining different approximations and approaches to the multi-stage optimization problem.

3.2.1 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 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:

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

This is a single-stage RO, with decision-variables $(\boldsymbol{x}_1, \boldsymbol{Q}, \boldsymbol{q})$. 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.

Despite this negative result, there are some positive complexity results concerning the affine model. In order to present these, we 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: $\boldsymbol{x}_2 = \boldsymbol{Q}\boldsymbol{u} + \boldsymbol{q}_0 = \boldsymbol{q}^{(0)} + \sum_{l=1}^{L} u_l \boldsymbol{q}^{(l)}$. 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

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

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 $\boldsymbol{u}$. The 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:

$$\alpha_i(\boldsymbol{\chi}) + 2\boldsymbol{u}^\top \beta_i(\boldsymbol{\chi}) + \boldsymbol{u}^\top \Gamma_i(\boldsymbol{\chi}) \boldsymbol{u} \geq 0, \quad \forall \boldsymbol{u} \in \mathcal{U}. \tag{3.9}$$

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:

$$\begin{aligned}
\min: \quad & \boldsymbol{c}^\top \boldsymbol{x}_1 \\
\text{s.t.}: \quad & \left(\begin{array}{c|c} \Gamma_i(\boldsymbol{\chi}) + \rho^{-2} \sum_{k=1}^K \lambda_k S_k & \beta_i(\boldsymbol{\chi}) \\ \hline \beta_i(\boldsymbol{\chi})^\top & \alpha_i(\boldsymbol{\chi}) - \sum_{k=1}^K \lambda_k^{(i)} \end{array}\right) \succeq \boldsymbol{0}, \ i = 1, \ldots, m \\
& \lambda^{(i)} \geq 0, \ i = 1, \ldots, m
\end{aligned} \tag{3.10}$$

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

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

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 [46].

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

1. Integrated circuit design: In [73], 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. Portfolio Management: In [37], a two-stage portfolio allocation problem is considered. While the uncertainty model is data-driven, the basic framework for handling the multi-stage decision-making is based on RO techniques.

3. 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.

4. Network flows and Traffic Management: In [80], 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 [76], the authors consider a similar problem under transportation cost and demand uncertainty, extending the work in [80].

5. Chance constraints: In [42], the authors apply a modified model of affine adaptability to the stochastic programming setting, and show how this can improve approximations of so-called chance constraints. In [49], 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).

Additional work in affine adaptability has been done in [42], 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.

3.2.2 Discrete Variables

Consider now a multi-stage optimization where the future stage decisions are subject to integer constraints. The framework introduced above cannot address such a setup, since the second stage policies, $\boldsymbol{x}_2(\boldsymbol{u})$, are necessarily continuous functions of the uncertainty.

3.2.3 Finite Adaptability

The framework of Finite Adaptability, introduced in Bertsimas and Caramanis [22] and Caramanis [39], is designed to deal exactly with this setup. There, the second-stage variables, $\boldsymbol{x}(\boldsymbol{u})$, are piecewise constant functions of the uncertainty, with k pieces. 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 [72]). Consider a two-stage problem of the form

$$\begin{array}{rl} \min: & \boldsymbol{c}^\top \boldsymbol{x}_1 + \boldsymbol{d}^\top \boldsymbol{x}_2(\boldsymbol{u}) \\ \text{s.t.}: & \boldsymbol{A}_1(\boldsymbol{u}) + \boldsymbol{A}_2(\boldsymbol{u})\boldsymbol{x}_2(\boldsymbol{u}) \geq \boldsymbol{b}, \quad \forall \boldsymbol{u} \in \mathcal{U} \\ & \boldsymbol{x}_1 \in \mathcal{X}_1, \boldsymbol{x}_2 \in \mathcal{X}_2, \end{array} \tag{3.11}$$

where $\mathcal{X}_2$ may contain integrality constraints. In the finite adaptability framework, with k-piecewise constant second stage variables, this becomes

$$\text{Adapt}_k(\mathcal{U}) = \min_{\mathcal{U}=\mathcal{U}_1\cup\cdots\cup\mathcal{U}_k} \left[\begin{array}{rl} \min: & \boldsymbol{c}^\top \boldsymbol{x}_1 + \max\{\boldsymbol{d}^\top \boldsymbol{x}_2^{(1)}, \ldots, \boldsymbol{d}^\top \boldsymbol{x}_2^{(k)}\} \\ \text{s.t.}: & \boldsymbol{A}_1(\boldsymbol{u})\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{u})\boldsymbol{x}_2^{(1)} \geq \boldsymbol{b}, \quad \forall \boldsymbol{u} \in \mathcal{U}_1 \\ & \vdots \\ & \boldsymbol{A}_1(\boldsymbol{u})\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{u})\boldsymbol{x}_2^{(k)} \geq \boldsymbol{b}, \quad \forall \boldsymbol{u} \in \mathcal{U}_k \\ & \boldsymbol{x}_1 \in \mathcal{X}_1, \boldsymbol{x}_2^{(j)} \in \mathcal{X}_2. \end{array} \right].$$

If the partition of the uncertainty set, $\mathcal{U} = \mathcal{U}_1 \cup \cdots \cup \mathcal{U}_k$ 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. Furthermore, the static problem (i.e., with no adaptability) corresponds to the case $k = 1$, and hence if this is feasible, then the k-adaptable problem is feasible for any k. This allows the decision-maker to choose the appropriate level of adaptability. This flexibility may be particularly important for very large scale problems, where the nominal formulation is already on the border of what is currently tractable. We provide such an example, in an application of finite adaptability to Air Traffic Control below.

The complexity of finite adaptability is in finding a good partition of the uncertainty. Indeed, in general, computing the optimal partition even into two regions is NP-hard ([22],[39]). However, we also have the following positive complexity result. It says that if any one of the three quantities: (a) Dimension of the uncertainty; (b) Dimension of the decision-space; and (c) Number of uncertain constraints, is small, then computing the optimal 2-piecewise constant second stage policy can be done efficiently.

Theorem 8 ([22],[39]). *Consider a two-stage problem of the form in (3.11). Suppose the uncertainty set $\mathcal{U}$ is given as the convex hull N points. Let $d = \min(N, \dim\mathcal{U})$, let n be the dimension of the second-stage decision-variable, and m the number of uncertain constraints (the number of rows of $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$. Then the optimal hyperplane partition of $\mathcal{U}$ can be obtained in time exponential in $\min(d, n, m)$, and in particular, if the dimension of the problem, or the dimension of the decision-variables, or the number of uncertain constraints is small, then the 2-adaptable problem is tractable.*

This result is particularly pertinent for the framework of finite adaptability. In particular, consider the dimension of the uncertainty set. If $\mathcal{U}$ is truly high-dimensional, then a piecewise-constant second-stage policy with only a few pieces, would most likely not be effective. The application to Air Traffic Control ([39]) which we present below, gives an example where the dimension of the uncertainty is large, but can be approximated by a low-dimensional set, thus rendering finite adaptability an appropriate framework.

3.2.4 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 9 ([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 $\zeta_S \stackrel{\triangle}{=} \max\{\boldsymbol{d}(S) \, : \, \boldsymbol{d} \in \mathcal{U}\}$.*

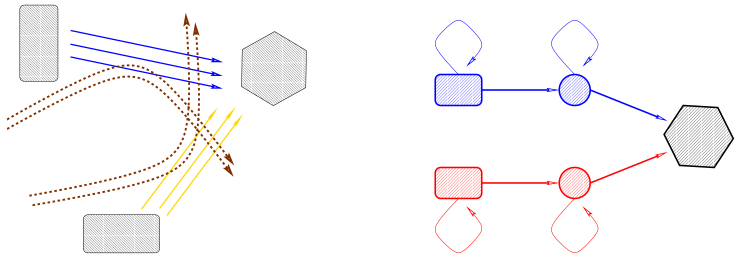

Figure 2: In the figure on the left, we have planes arriving at a single hub such as JFK in NYC. Dashed lines express uncertainty in the weather. The figure on the right gives the simplified version for the scenario we consider.

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 10 ([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.

3.2.5 Nonlinear Adaptability

There has also been some work on adaptability for nonlinear problems, in Takeda, Taguchi and Tütüncü [93]. General single-stage robustness is typically intractable. Thus one cannot expect far-reaching tractability results for the multi-stage case. Nevertheless, in this paper the authors offer sufficient conditions on the uncertainty set and the structure of the problem, so that the resulting nonlinear multi-stage robust problem is tractable. In [93], they consider several applications to portfolio management.

3.3 An Application of Robust Adaptable Optimization: Air Traffic Control

The 30,000 daily flights over the US Air Space (NAS) must be scheduled to minimize delay, while respecting the weather impacted, and hence uncertain, takeoff, landing, and in-air capacity constraints. Because of the discrete variables, continuous adaptability cannot work. Also, because of the large-scale nature of the problem, there is very little leeway to increase the size of the problem. We give a small example (see [39] for more details and computations) to illustrate the application of Finite Adaptability.

Figure 1 depicts a major airport (e.g., JFK) that accepts heavy traffic from airports to the West and the South. In this figure, the weather forecast predicts major local disruption due to an approaching storm, affecting only the immediate vicinity of the airport; the timing of the impact, however, is uncertain, and at question is which of the 50 (say) northbound and 50 eastbound flights to hold on the ground, and which to hold in the air. We assume the direct (undelayed) flight time is 2 hours. Each plane

	Delay Cost	Ground Holding	Air Holding
Utopic:	2,050	205	0
Static:	4,000	400	0
2-Adaptable:	3,300	170	80
4-Adaptable:	2,900	130	80

Table 2: Results for the delay costs for the utopic, robust, 2-adaptable, and 4-adaptable schemes.

may be held either on the ground, in the air, or both, for a total delay not exceeding 60 minutes. The simplified picture is presented in Figure 1 on the right. Rectangular nodes represent the airports, and the self-link ground holding. The intermediate circular nodes represent a location one hour from JFK, in a geographical region whose capacity is unaffected by the storm. The self-link here represents air holding. The final hexagonal node represents the destination airport, JFK. Thus the links from the two circular nodes to the final hexagonal node are the only capacitated links in this simple example.

We discretize time into 10-minute intervals. We assume that the impact of the storm lasts 30 minutes, with the timing and exact directional approach uncertain. Because we are discretizing time into 10 minute intervals, there are four possible realizations of the weather-impacted capacities in the second hour of our horizon. We give the capacity in terms of the number of planes per 10-minute interval:

$$(1)\ \left[\begin{array}{lrrrrrr} \text{West:} & 15 & 15 & 15 & \underline{5} & \underline{5} & \underline{5} \\ \text{South:} & \underline{5} & \underline{5} & \underline{5} & 15 & 15 & 15 \end{array}\right] \quad (2)\ \left[\begin{array}{lrrrrrr} \text{West:} & 15 & 15 & \underline{5} & \underline{5} & \underline{5} & 15 \\ \text{South:} & 15 & \underline{5} & \underline{5} & \underline{5} & 15 & 15 \end{array}\right]$$

$$(3)\ \left[\begin{array}{lrrrrrr} \text{West:} & 15 & \underline{5} & \underline{5} & \underline{5} & 15 & 15 \\ \text{South:} & 15 & 15 & \underline{5} & \underline{5} & \underline{5} & 15 \end{array}\right] \quad (4)\ \left[\begin{array}{lrrrrrr} \text{West:} & \underline{5} & \underline{5} & \underline{5} & 15 & 15 & 15 \\ \text{South:} & 15 & 15 & 15 & \underline{5} & \underline{5} & \underline{5} \end{array}\right]$$

In the utopic set-up (not implementable) the decision-maker can foresee the future (of the storm) and makes decisions accordingly. Thus we get a bound on performance. We also consider a nominal, no-robustness scheme, where the decision-maker (naïvely) assumes the storm will behave exactly according to the first scenario. We also consider adaptabiliy formulations: 1-adaptable (static robust) solution, then the 2- and 4-adaptable solution. Each 10-minute interval of ground delay adds 10 to the cost, while air-delay adds 20 (per flight).

4 Applications of Robust Optimization

In this section, we survey the main applications modeled by Robust Optimization techniques.

4.1 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. We now describe some of the noteworthy approaches and results in more detail.

4.1.1 Uncertainty models for return mean and covariance

The classical work of Markowitz ([74, 75]) 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., $\min\{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \ : \ \boldsymbol{\mu}^\top \boldsymbol{w} \geq r, \ \boldsymbol{w} \in \mathcal{W}\}$ or, alternatively, the *maximum return problem*, i.e., $\min\{\boldsymbol{\mu}^\top \boldsymbol{w} \ : \ \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \leq \sigma^2, \ \boldsymbol{w} \in \mathcal{W}\}$. 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).

Despite the widespread popularity of this approach, 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 [70] propose box, ellipsoidal, and other uncertainty sets for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. With these uncertainty structures, they provide a polynomial-time cutting plane algorithm for solving robust variants, e.g., the *robust minimum variance problem*

$$\begin{array}{ll} \min\limits_{\boldsymbol{w} \in \mathcal{W}} & \sup\limits_{\boldsymbol{\Sigma} \in \mathcal{S}} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \\ \text{subject to} & \inf\limits_{\boldsymbol{\mu} \in \mathcal{M}} \boldsymbol{\mu}^\top \boldsymbol{w} \geq r. \end{array} \tag{4.12}$$

Costa and Paiva [43] 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 the portfolio problems as optimization problems over linear matrix inequalities.

Tütüncü and Koenig [94] focus on the case of box uncertainty sets for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as well and show that Problem (4.12) is equivalent to the *robust risk-adjusted return problem*

$$\min_{\boldsymbol{w} \in \mathcal{W}} \quad \inf_{\boldsymbol{\mu} \in \mathcal{M}, \ \boldsymbol{\Sigma} \in \mathcal{S}} \left\{ \boldsymbol{\mu}^\top \boldsymbol{w} - \lambda \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \right\},$$

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 [60] to compute robust efficient frontiers.

4.1.2 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. [57], 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 defined as $\mathrm{VaR}_\epsilon(\boldsymbol{w}) \triangleq \min\{\gamma \ : \ \mathbb{P}\left(\gamma \leq -\tilde{\boldsymbol{r}}^\top \boldsymbol{w}\right) \leq \epsilon\}$. In turn, the authors in [57] approach the worst-case VaR problem, i.e.,

$$\min_{\boldsymbol{w}\in\mathcal{W}} \quad V_{\mathcal{P}}(\boldsymbol{w}), \tag{4.13}$$

where $V_{\mathcal{P}}(\boldsymbol{w}) \triangleq \min\{\gamma \ : \ \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\gamma \leq -\tilde{\boldsymbol{r}}^\top \boldsymbol{w}\right) \leq \epsilon\}$. In particular, the authors first focus on the distributional family $\mathcal{P}$ with fixed mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma} \succ \boldsymbol{0}$. From a tight Chebyshev bound (e.g., Bertsimas and Popescu [24]), it is known that (4.13) is equivalent to the SOCP $\min\{\gamma \ : \ \kappa(\epsilon)\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\|_2 - \boldsymbol{\mu}^\top\boldsymbol{w} \leq \gamma\}$, where $\kappa(\epsilon) = \sqrt{(1-\epsilon)/\epsilon}$; in [57], however, the authors also show equivalence of (4.13) to an SDP, and this allows them to extend to the case of uncertainty in the moment information. Specifically, when the supremum in (4.13) is taken over all distributions with mean and covariance known only to belong within $\mathcal{U}$, i.e., $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathcal{U}$, [57] shows the following:

1. When $\mathcal{U} = \operatorname{conv}\{(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \ldots, (\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)\}$, then (4.13) is SOCP-representable.
2. When $\mathcal{U}$ is a set of component-wise box constraints on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, then (4.13) is SDP-representable.

One interesting extension in [57] 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 $\int \log \frac{d\mathbb{P}}{d\mathbb{P}_0} d\mathbb{P} \leq d$ in (4.13) still leads to an SOCP-representable problem, with $\kappa(\epsilon)$ modified to a new value $\kappa(\epsilon, d)$. Thus, imposing this closeness condition on the distributional family only requires modification of the risk factor.

Pinar and Tütüncü [86] 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 $\boldsymbol{\mu}^\top\boldsymbol{w} - \theta\sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}} \geq 0$, and $\boldsymbol{e}^\top\boldsymbol{x} < 0$. The rationale for this is the fact shown by Ben-Tal and Nemirovski [15] that the probability that a bounded random variable is less than θ standard deviations below its mean is less than $e^{-\theta^2/2}$. Therefore, θ-robust profit portfolios return a positive amount with very high probability. The authors in [86] then attempt to solve the *maximum-θ robust profit opportunity problem*:

$$\begin{aligned} \sup_{\theta, \boldsymbol{w}} \quad & \theta \\ \text{subject to} \quad & \boldsymbol{\mu}^\top\boldsymbol{w} - \theta\sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}} \geq 0 \\ & \boldsymbol{e}^\top\boldsymbol{w} < 0, \end{aligned} \tag{4.14}$$

and show that (4.14) is equivalent to a convex quadratic program and derive closed-form solutions under mild conditions. Moreover, when there is also a risk-free asset, maximum-θ robust profit portfolios are maximum Sharpe ratio [90] portfolios.

4.1.3 Robust factor models

A common practice in modeling market return dynamics is to use a so-called *factor model* of the form $\tilde{\boldsymbol{r}} = \boldsymbol{\mu} + \boldsymbol{V}^\top \boldsymbol{f} + \boldsymbol{\epsilon}$, 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 (typically major stock indices or other 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 this have been considered by a few authors. Goldfarb and Iyengar [59] use the following uncertainty model for the parameters

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

where $\boldsymbol{f} \in \mathcal{N}(\boldsymbol{0}, \boldsymbol{F})$, $\boldsymbol{\epsilon} \in \mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$, $\boldsymbol{W}_i = \boldsymbol{W}\boldsymbol{e}_i$ 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, 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, under a particular ellipsoidal uncertainty model the factor covariance matrix $\boldsymbol{F}$ can be included in the robust problems and the resulting problem may still be formulated as an SOCP.

In [57], the authors show how to compute upper bounds on the robust worst-case VaR problem with a factor model via SDP for joint ellipsoidal and norm-bounded uncertainty models in $(\boldsymbol{\mu}, \boldsymbol{V})$.

4.1.4 Multi-period robust models

The robust portfolio models discussed heretofore have been for single-stage problems. Some efforts have been made on multi-stage problems. Especially notable is the work of Ben-Tal et al. [11], who formulate the following, L-stage portfolio problem:

$$
\begin{array}{ll}
\text{maximize} & \displaystyle\sum_{i=1}^{n+1} r_i^L x_i^L \\
\text{subject to} & x_i^l = r_i^{l-1} x_i^{l-1} - y_i^l + z_i^l,\ i = 1, \ldots, n,\ l = 1, \ldots, L \\
& x_{n+1}^l = r_{n+1}^{l-1} x_{n+1}^{l-1} + \displaystyle\sum_{i=1}^{n} (1 - \mu_i^l) y_i^l - \sum_{i=1}^{n} (1 + \nu_i^l) z_i^l,\ l = 1, \ldots, L \\
& x_i^l, y_i^l, z_i^l \geq 0,
\end{array}
\tag{4.15}
$$

where 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, (4.15) as stated is simply a linear programming problem and contains no reference to the uncertainty in the returns and the transaction costs. One can utilize a multi-stage stochastic programming approach to the problem, but this is extremely onerous 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 [86]) for the uncertain parameters, and show how to solve a "rolling horizon" version of the problem via SOCP.

From a structural standpoint, the authors in [11] are also able to show that solutions to their robust version of (4.15) obey the property that one never both buys and sells an asset i during a single time period l for all asset/time index pairs (i, l) satisfying a specific second moment condition on the uncertainties. In these cases, the robust version of (4.15) matches the intuition that, because of transaction costs, one should never both buy and sell an asset simultaneously.

Pinar and Tütüncü [86] explore a two-period model for their robust profit opportunity problem. In particular, they examine the problem

$$\begin{array}{ll} \sup\limits_{\boldsymbol{x}_0} & \inf\limits_{\boldsymbol{r}^1 \in \mathcal{U}} \sup\limits_{\theta, \boldsymbol{x}^1} \theta \\ \text{subject to} & \boldsymbol{e}^\top \boldsymbol{x}^1 = (\boldsymbol{r}^1)^\top \boldsymbol{x}^0 \qquad \text{(self-financing constraint)} \\ & (\boldsymbol{\mu}^2)^\top \boldsymbol{x}^1 - \theta \sqrt{(\boldsymbol{x}^1)^\top \boldsymbol{\Sigma}_2 \boldsymbol{x}^1} \geq 0 \\ & \boldsymbol{e}^\top \boldsymbol{x}^0 < 0, \end{array} \tag{4.16}$$

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 (4.16) depends critically on $\mathcal{U}$, but [86] derives a solution to the problem when $\mathcal{U}$ is ellipsoidal.

4.1.5 Computational results for robust portfolios

Most of the studies on robust portfolio optimization are corroborated by promising computational experiments. Here we provide a short though by no means exhaustive summary of such results.

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 which uses expected values of the return vectors.

Goldfarb and Iyengar [59] 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. [57] 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 [94] 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. [48] 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 [59] 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.

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.

4.2 Statistics, learning, and estimation

The process of using data to analyze or describe the parameters and behavior of a system is inherently uncertain, and RO has been applied in many contexts. We now touch upon some of these.

4.2.1 Least-squares problems

The problem of robust, least-squares solutions to systems of over-determined linear equations is considered by El Ghaoui and Lebret [56]. Specifically, 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 [56], the authors build explicit models to account for uncertainty for the data $[\boldsymbol{A}\ \boldsymbol{b}]$. Prior to this work, there existed numerous regularization techniques for handling this uncertainty, but no explicit, robust models. The authors consider the *Robust Least-Squares (RLS) Problem*:

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

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

[56] then shows that RLS may be formulated as an SOCP, which, in turn, may be further reduced to a one-dimensional convex optimization problem. Moreover, the authors show that there exists a threshold uncertainty level $\rho_{\min}(\boldsymbol{A}, \boldsymbol{b})$ (which is explicitly computed) such that, for all $\rho \leq \rho_{\min}(\boldsymbol{A}, \boldsymbol{b})$, the solutions to the ordinary least-squares and RLS coincide. Thus, ordinary least-squares solutions are $\rho_{\min}(\boldsymbol{A}, \boldsymbol{b})$-robust.

4.2.2 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 $\mathcal{H}(\boldsymbol{a}, b) = \left\{\boldsymbol{z} \in \mathbb{R}^n \mid \boldsymbol{a}^\top \boldsymbol{z} = b\right\}$, 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.

Lanckriet et al. [69] 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. In particular, the *Minimax Probability Machine* (MPM) finds a separating hyperplane $(\boldsymbol{a}, b)$ to the problem

$$\begin{array}{ll} \text{maximize} & \alpha \\ \text{subject to} & \inf\limits_{\boldsymbol{x} \sim (\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)} \mathbb{P}\left(\boldsymbol{a}^\top \boldsymbol{x} \geq b\right) \geq \alpha \\ & \inf\limits_{\boldsymbol{y} \sim (\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)} \mathbb{P}\left(\boldsymbol{a}^\top \boldsymbol{y} \leq b\right) \geq \alpha, \end{array} \tag{4.17}$$

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 (4.17) can be solved via SOCP, and the worst-case misclassification probability is given as $1/(1+\kappa_*^2)$, where κ_*^{-1} is the optimal value of the SOCP formulation. They then proceed to enhance the model by accounting for uncertainty in the means and covariances. The robust problem in this case is the same as (4.17) but the constraints must hold for all $(\boldsymbol{\mu}_x \boldsymbol{\Sigma}_x) \in \mathcal{X}$, $(\boldsymbol{\mu}_y \boldsymbol{\Sigma}_y) \in \mathcal{Y}$, with the following uncertainty model for the means and covariances considered:

$$\begin{aligned} \mathcal{X} &= \left\{(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \mid (\boldsymbol{\mu}_x - \boldsymbol{\mu}_x^0)^\top \boldsymbol{\Sigma}_x^{-1} (\boldsymbol{\mu}_x - \boldsymbol{\mu}_x^0) \leq \nu^2,\ \|\boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_x^0\|_F \leq \rho\right\}, \\ \mathcal{Y} &= \left\{(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y) \mid (\boldsymbol{\mu}_y - \boldsymbol{\mu}_y^0)^\top \boldsymbol{\Sigma}_y^{-1} (\boldsymbol{\mu}_y - \boldsymbol{\mu}_y^0) \leq \nu^2,\ \|\boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_y^0\|_F \leq \rho\right\}. \end{aligned}$$

The authors in [69] show that this robust version is equivalent to an appropriately defined, nominal MPM problem of the form (4.17), in particular the one with $\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_X^0 + \rho \boldsymbol{I}$ and $\boldsymbol{\Sigma}_y = \boldsymbol{\Sigma}_y^0 + \rho \boldsymbol{I}$. In addition, the worst-case misclassification probability of the robust version is $1/(1 + \max(0, \kappa_* - \nu)^2)$.

El Ghaoui [55] et al. consider binary classification problems using an uncertainty model on the observations directly. The notation used is slightly different. Here, let $\boldsymbol{X} \in \mathbb{R}^{n \times N}$ be a matrix with the N columns each corresponding to an observation, and let $\boldsymbol{y} \in \{-1, +1\}^n$ be an associated label vector denoting class membership. [55] considers an interval uncertainty model for $\boldsymbol{X}$:

$$\mathcal{X}(\rho) = \left\{\boldsymbol{Z} \in \mathbb{R}^{n \times N} \mid \boldsymbol{X} - \rho \boldsymbol{\Sigma} \leq \boldsymbol{Z} \leq \boldsymbol{X} + \rho \Sigma\right\}, \tag{4.18}$$

where $\boldsymbol{\Sigma}$ and $\rho \geq 0$ are pre-specified parameters. They then seek a linear classification rule based on the sign of $\boldsymbol{a}^\top \boldsymbol{x} - b$, where $\boldsymbol{a} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ and $b \in \mathbb{R}$ are decision variables. The *robust classification problem with interval uncertainty* is

$$\min_{\boldsymbol{a} \neq \boldsymbol{0}, b} \quad \max_{\boldsymbol{Z} \in \mathcal{X}(\rho)} L(\boldsymbol{a}, b, \boldsymbol{Z}, \boldsymbol{y}), \tag{4.19}$$

where L is a particular *loss function.* The authors then compute explicit, convex optimization problems for several types of commonly used loss functions (support vector machines, logistic regression, and minimax probability machines; see [55] for the full details).

Another technique for linear classification is based on so-called *Fisher discriminant analysis* (FDA) [51]. 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

$$f(\boldsymbol{a}, \boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_y) \quad := \quad \frac{\left(\boldsymbol{a}^\top(\boldsymbol{\mu}_x - \boldsymbol{\mu}_y)\right)^2}{\boldsymbol{a}^\top\left(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y\right)\boldsymbol{a}}, \tag{4.20}$$

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 $\boldsymbol{a}^{\text{nom}} = (\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y)^{-1}(\boldsymbol{\mu}_x - \boldsymbol{\mu}_y)$ gives the maximum value of this ratio. Kim et al. [67] consider the *robust Fisher linear discriminant problem*

$$\text{maximize}_{\boldsymbol{a}\neq\boldsymbol{0}} \quad \min_{(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_y)\in\mathcal{U}} f(\boldsymbol{a}, \boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_y), \tag{4.21}$$

where $\mathcal{U}$ is any convex uncertainty set for the mean and covariance parameters. [67] then shows that the discriminant $\boldsymbol{a}^* \triangleq \left(\boldsymbol{\Sigma}_x^* + \boldsymbol{\Sigma}_y^*\right)^{-1}(\boldsymbol{\mu}_x^* - \boldsymbol{\mu}_y^*)$ is optimal to the Robust Fisher linear discriminant problem (4.21), where $(\boldsymbol{\mu}_x^*, \boldsymbol{\mu}_y^*, \boldsymbol{\Sigma}_x^*, \boldsymbol{\Sigma}_y^*)$ is any optimal solution to the convex optimization problem:

$$\min_{(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_y)\in\mathcal{U}} \quad (\boldsymbol{\mu}_x - \boldsymbol{\mu}_y)^\top(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y)^{-1}(\boldsymbol{\mu}_x - \boldsymbol{\mu}_y).$$

Other work using robust optimization for classification and learning, includes that of Shivaswamy et al. [91] who consider SOCP approaches for handling missing and uncertain data, and also Caramanis and Mannor [40], where robust optimization is used to obtain a model for uncertainty *in the label* of the training data.

4.2.3 Parameter estimation

Calafiore and El Ghaoui [38] 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 $\bar{\boldsymbol{y}} = \boldsymbol{C}(\boldsymbol{\Delta}_c)\bar{\boldsymbol{x}}$. 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 $\bar{\boldsymbol{x}}_{\text{ML}}(\boldsymbol{\Delta}) \triangleq \arg\min_{\bar{\boldsymbol{x}}} \|F(\boldsymbol{\Delta})\bar{\boldsymbol{x}} - g(\boldsymbol{\Delta})\|^2$, where $\boldsymbol{\Delta} = [\boldsymbol{\Delta}_p^\top \ \boldsymbol{\Delta}_d^\top \ \boldsymbol{\Delta}_c^\top]$ and $F(\boldsymbol{\Delta})$ and $g(\boldsymbol{\Delta})$ are functions of $\boldsymbol{D}(\boldsymbol{\Delta}_d))$, $\boldsymbol{P}(\boldsymbol{\Delta}_p))$, and $\boldsymbol{C}(\boldsymbol{\Delta}_c))$.

The authors in [38] 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 $\boldsymbol{\Delta}_1 \triangleq \left\{\boldsymbol{\Delta} \in \hat{\boldsymbol{\Delta}} \;\middle|\; \|\boldsymbol{\Delta}\| \leq 1\right\}$, for $\hat{\boldsymbol{\Delta}}$ a linear subspace (e.g., $\mathbb{R}^{p\times q}$) and $||\cdot||$ the spectral (maximum singular value) norm. The robust or *worst-case maximum likelihood* (WCML) problem, then, is

$$\text{minimize} \quad \max_{\boldsymbol{\Delta}\in\boldsymbol{\Delta}_1} \|F(\boldsymbol{\Delta})\boldsymbol{x} - g(\boldsymbol{\Delta})\|^2. \tag{4.22}$$

The work [38] shows that the WCML problem (4.22) 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. [47] 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 $\min_{\hat{\boldsymbol{x}}} \mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2\right]$. 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 $\|\boldsymbol{x}\|_{\boldsymbol{T}} \leq L$, where $\|\boldsymbol{x}\|_{\boldsymbol{T}}^2 \triangleq \boldsymbol{x}^\top \boldsymbol{T}\boldsymbol{x}$ for some known, positive definite matrix $\boldsymbol{T} \in \mathbb{S}^n$, and $L \geq 0$. The *worst-case MSE problem* then is

$$\min_{\hat{\boldsymbol{x}}=\boldsymbol{G}\boldsymbol{y}} \max_{\{\|\boldsymbol{x}\|_{\boldsymbol{T}} \leq L\}} \mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2\right]. \tag{4.23}$$

Notice that this problem restricts to estimators which are linear in the observations. [47] then shows that (4.23) 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 $\boldsymbol{H}$, which they also show is an SDP.

4.3 Supply chain management

Bertsimas and Thiele [28] consider a robust model for inventory control as discussed above in Section **??**. They use a cardinality-constrained uncertainty set, as developed in Section 2.2. One main contribution of [28] 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 [28] 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 is as previously discussed in Section 3. This model has adaptability explicitly integrated into it, but computed as an *affine* function of the realized demands. 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 [30] 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.

4.4 Engineering

Robust Optimization techniques have been applied to a wide variety of engineering problems. In this section, we briefly mention some of the work in this area. We omit most technical details and refer the reader to the relevant papers for more. Some of the many engineering applications are as follows.

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.

Circuit design: Boyd et al. [33] and Patil et al. [85] 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 [73] and [72], already discussed above.

Power control in wireless channels: Hsiung et al. [62] 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.

Antenna design: Lorenz and Boyd [71] 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. [77] consider beamforming design where the weights cannot be implemented exactly, but instead are only known to lie within a box constraint. They show that the resulting design problem has the same structure as the nominal beamforming problem and may, in fact, be interpreted as a regularized version of this nominal problem.

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. [95]). Somewhat in contrast to this literature, Bertsimas and Brown [21] explicitly use recent RO techniques to develop a tractable approach to constrained linear-quadratic control problems.

References

[1] E. Adida and G. Perakis. A robust optimization approach to dynamic pricing and inventory control with no backorders. *Mathematical Programming*, 107(1):97–129, 2006.

[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. Forthcoming in *Mathematical Programming*, 2005.

[4] A. Atamtürk and M. Zhang. Two-stage robust network flow and design under demand uncertainty. Technical Report BCOL.04.03, IEOR, University of California–Berkeley, December 2004.

[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 flexible approach to robust optimization via convex risk measures. Submitted, 2006.

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

[20] D. Bertsimas and D.B. Brown. Constructing uncertainty sets for robust linear optimization. Submitted, 2006.

[21] D. Bertsimas and D.B. Brown. Constrained Stochastic LQC: A Tractable Approach. To appear in *IEEE Transactions on Automatic Control*, 2007.

[22] D. Bertsimas and C. Caramanis. Finite adaptability in linear optimization. Technical Report available from: `http://users.ece.utexas.edu/∼cmcaram/`, The University of Texas at Austin, February 2007.

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

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

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

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

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

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

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

[30] D. Bienstock and N. Özbay. Computing robust basestock levels. Working paper, 2006.

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

[32] 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.

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

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

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

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

[37] 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.

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

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

[40] C. Caramanis and S. Mannor. Beyond PAC: A robust optimization approach for learning in the presence of noise. Technical Report available from: `http://users.ece.utexas.edu/ cmcaram`, The University of Texas at Austin, 2007.

[41] X. Chen, M. Sim, and P. Sun. A robust optimization perspective to stochastic programming. To appear in *Operations Research*, 2007.

[42] X. Chen, M. Sim, P. Sun, and J. Zhang. A linear-decision based approximation approach to stochastic programming. To appear in *Operations Research*, 2007.

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

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

[45] E. Delage and S. Mannor. Percentile optimization for markov decision processes with parameter uncertainty. Submitted to *Operations Research.*

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

[47] 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.

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

[49] E. Erdoğan and G. Iyengar. On two-stage convex chance constrained problems. Technical Report CORC TR-2005-06, IEOR, Columbia University, November 2005.

[50] J.E. Falk. Exact solutions of inexact linear programs. *Operations Research*, 24:783–787, 1976.

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

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

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

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

[55] L. El Ghaoui, G.R.G. Lanckriet, and G. Natsoulis. Robust classification with interval data. Report No. UCB/CSD-03-1279, October, 2003.

[56] 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.

[57] 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.

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

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

[60] 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.

[61] K.-L. Hsiung, S.-J. Kim, and S. Boyd. Tractable approximate robust geometric programming. Submitted to *Mathematical Programming*, 2005.

[62] 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.

[63] X. Huan and S. Mannor. Tradeoff performance and robustness in linear programming and markov decision processes. In *NIPS*, 2006.

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

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

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

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

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

[69] 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.

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

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

[72] M. Mani, C. Caramanis, and M. Orshansky. Power efficient buffer design using finite adaptable optimization. In *Submitted to ICCAD*, 2007.

[73] 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.*

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

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

[76] 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.

[77] A. Mutapcic, S.-J. Kim, and S. Boyd. Beamforming with uncertain weights. To appear in *IEEE Signal Processing Letters*, 2007.

[78] K. Natarajan, D. Pachamanova, and M. Sim. Constructing risk measures from uncertainty sets. Working paper, 2005.

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

[80] F. Ordó nez and J. Zhao. Robust capacity expansion of network flows. Technical report, University of Southern California, May 2005.

[81] A. Nilim and L. El Ghaoui. Air traffic control under stochastic environments. In *American Control Conference*, 2004.

[82] A. Nilim and L. El Ghaoui. Robust markov decision processes with uncertain transition matrices. Technical Report M04-26, Berkeley EECS, January 2004.

[83] A. Nilim, L. El Ghaoui, M. Hansen, and V. Duong. Trajectory-based air traffic management under weather uncertainty. In *Fourth USA/EUROPE ATM R & D Seminar*, 2001.

[84] 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.

[85] 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.

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

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

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

[89] A. Shapiro. On complexity of multistage stochastic programs. *Optimization Online*, 2005.

[90] W. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964.

[91] 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.

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

[93] A. Takeda, S. Taguchi, and R. Tütüncü. Adjustable robust optimization models for nonlinear multi-period optimization. *Submitted*, 2004.

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

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