# Robust and Data-Driven Optimization: Modern Decision-Making Under Uncertainty

Dimtris Bertsimas[*] Aurélie Thiele[†]

March 2006

**Abstract**

Traditional models of decision-making under uncertainty assume perfect information, i.e., accurate values for the system parameters and specific probability distributions for the random variables. However, such precise knowledge is rarely available in practice, and a strategy based on erroneous inputs might be infeasible or exhibit poor performance when implemented. The purpose of this tutorial is to present a mathematical framework that is well-suited to the limited information available in real-life problems and captures the decision-maker's attitude towards uncertainty; the proposed approach builds upon recent developments in robust and data-driven optimization. In robust optimization, random variables are modeled as uncertain parameters belonging to a convex uncertainty set and the decision-maker protects the system against the worst case within that set. Data-driven optimization uses observations of the random variables as direct inputs to the mathematical programming problems. The first part of the tutorial describes the robust optimization paradigm in detail in single-stage and multi-stage problems. In the second part, we address the issue of constructing uncertainty sets using historical realizations of the random variables and investigate the connection between convex sets, in particular polyhedra, and a specific class of risk measures.

Keywords: optimization under uncertainty; risk preferences; uncertainty sets; linear programming.

## 1 Introduction

The field of decision-making under uncertainty was pioneered in the 1950s by Dantzig [25] and Charnes and Cooper [23], who set the foundation for, respectively, stochastic programming and

[*]Boeing Professor of Operations Research, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139, dbertsim@mit.edu

[†]P.C. Rossin Assistant Professor, Department of Industrial and Systems Engineering, Lehigh University, Mohler Building Room 329, Bethlehem, PA 18015, aurelie.thiele@lehigh.edu.

optimization under probabilistic constraints. While these classes of problems require very different models and solution techniques, they share the same assumption that the probability distributions of the random variables are known exactly, and despite Scarf's [38] early observation that "we may have reason to suspect that the future demand will come from a distribution that differs from that governing past history in an unpredictable way," the majority of the research efforts in decision-making under uncertainty over the past decades have relied on the precise knowledge of the underlying probabilities. Even under this simplifying assumption, a number of computational issues arises, e.g., the need for multi-variate integration to evaluate chance constraints and the large-scale nature of stochastic programming problems. The reader is referred to Birge and Louveaux [22] and Kall and Mayer [31] for an overview of solution techniques. Today, stochastic programming has established itself as a powerful modeling tool when an accurate probabilistic description of the randomness is available; however, in many real-life applications the decision-maker does not have this information, for instance when it comes to assessing customer demand for a product. (The lack of historical data for new items is an obvious challenge to estimating probabilities, but even well-established product lines can face sudden changes in demand, due to the market entry by a competitor or negative publicity.) Estimation errors have notoriously dire consequences in industries with long production lead times such as automotive, retail and high-tech, where they result in stockpiles of unneeded inventory or, at the other end of the spectrum, lost sales and customers' dissatisfaction. The need for an alternative, non-probabilistic, theory of decision-making under uncertainty has become pressing in recent years because of volatile customer tastes, technological innovation and reduced product life cycles, which reduce the amount of information available and make it obsolete more quickly.

In mathematical terms, imperfect information threatens the relevance of the solution obtained by the computer in two important aspects: (i) the solution might not actually be feasible when the decision-maker attempts to implement it, and (ii) the solution, when feasible, might lead to a far greater cost (or smaller revenue) than the truly optimal strategy. Potential infeasibility, e.g., from errors in estimating the problem parameters, is of course the primary concern of the decision-maker. The field of operations research remained essentially silent on that issue until Soyster's work [44], where every uncertain parameter in convex programming problems was taken equal to its worst-case value within a set. While this achieved the desired effect of immunizing the problem against parameter uncertainty, it was widely deemed too conservative for practical implementation. In the mid-1990s, research teams led by Ben-Tal and Nemirovski ([7], [8], [9]) and El-Ghaoui and Lebret ([27], [28]) addressed the issue of overconservatism by restricting the uncertain parameters to belong to ellipsoidal uncertainty sets, which removes the most unlikely outcomes from consideration and yields tractable mathematical programming problems. In line with these authors' terminology, optimization for the worst-case value of parameters within a set has become known as "robust optimization." A drawback of the robust modeling framework with ellipsoidal uncertainty sets is that it increases the complexity of the problem considered, e.g., the robust counterpart of a linear programming problem is a second-order cone problem. More recently, Bertsimas and Sim ([17], [18]) and Bertsimas et. al. [16] have proposed a robust optimization approach based on polyhedral uncertainty sets, which preserves the class of problems under analysis, e.g., the robust counterpart of a linear programming problem remains a linear programming problem, and thus has advantages in terms of tractability in large-scale settings. It can also be connected to the decision-maker's attitude towards uncertainty, providing guidelines to construct the uncertainty set from the historical realizations of the random variables using data-driven optimization (Bertsimas and Brown [13]).

The purpose of this tutorial is to illustrate the capabilities of the robust and data-driven optimization framework as a modeling tool in decision-making under uncertainty, and in particular to:

1. Address estimation errors of the problem parameters and model random variables in single-stage settings (Section 2),
2. Develop a tractable approach to dynamic decision-making under uncertainty, incorporating the fact that information is revealed in stages (Section 3),
3. Connect the decision-maker's risk preferences with the choice of uncertainty set using the available data (Section 4).

# 2 Static Decision-Making under Uncertainty

## 2.1 Uncertainty Model

In this section, we present the robust optimization framework when the decision-maker must select a strategy before (or without) knowing the exact value taken by the uncertain parameters. Uncertainty can take two forms: (i) estimation errors for parameters of constant but unknown value, and (ii) stochasticity of random variables. The model here does not allow for recourse, i.e, remedial action once the values of the random variables become known. Section 3 addresses the case where the decision-maker can adjust his strategy to the information being revealed over time.

Robust optimization builds upon the following two principles, which have been identified by Nahmias [32], Simchi-Levi et. al. [43] and Sheffi [41] as fundamental to the practice of modern operations management under uncertainty:

- Point forecasts are meaningless (because they are always wrong) and should be replaced by range forecasts.
- Aggregate forecasts are more accurate than individual ones.

The framework of robust optimization incorporates these managerial insights into quantitative decision models as follows. We model uncertain quantities (parameters or random variables) as parameters belonging to a prespecified interval, the *range forecast*, provided for instance by the marketing department. Such forecasts are in general symmetric around the point forecast, i.e., the nominal value of the parameter considered. The greater accuracy of aggregate forecasting will be incorporated by an *additional constraint* limiting the maximum deviation of the aggregate forecast from its nominal value.

To present the robust framework in mathematical terms, we follow closely Bertsimas and Sim [18] and consider the linear programming problem:

$$\begin{aligned} \min \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b}, \\ & \mathbf{x} \in X, \end{aligned} \tag{1}$$

where uncertainty is assumed without loss of generality to affect only the constraint coefficients $\mathbf{A}$ and $X$ is a polyhedron (not subject to uncertainty). Problem (1) arises in a wide range of settings; it can for instance be interpreted as a *production planning problem* where the decision-maker must purchase raw material to minimize cost while meeting the demand for each product, despite uncertainty on the machine productivities. Note that a problem with uncertainty in the cost vector $\mathbf{c}$ and the right-hand side $\mathbf{b}$ can immediately be reformulated as:

$$\begin{aligned} \min \quad & Z \\ \text{s.t.} \quad & Z - \mathbf{c}'\mathbf{x} \geq 0, \\ & \mathbf{A}\mathbf{x} - \mathbf{b}\, y \geq \mathbf{0}, \\ & \mathbf{x} \in X,\ y = 1, \end{aligned} \tag{2}$$

which has the form of Problem (1).

The fundamental issue in Problem (1) is one of *feasibility*; in particular, the decision-maker will guarantee that every constraint is satisfied for any possible value of $\mathbf{A}$ in a given convex uncertainty set $\mathcal{A}$ (which will be described in detail shortly). This leads to the following formulation of the

**robust counterpart** of Problem (1):

$$\begin{aligned} \min \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{a_i}'\mathbf{x} \geq b_i, \quad \forall i, \forall \mathbf{a_i} \in \mathcal{A}, \\ & \mathbf{x} \in X, \end{aligned} \tag{3}$$

or equivalently:

$$\begin{aligned} \min \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \min_{\mathbf{a_i} \in \mathcal{A}} \mathbf{a_i}'\mathbf{x} \geq b_i, \quad \forall i, \\ & \mathbf{x} \in X, \end{aligned} \tag{4}$$

where $\mathbf{a_i}$ is the $i$-th vector of $\mathbf{A}'$.

Solving the robust problem as it is formulated in Problem (4) would require evaluating $\min_{\mathbf{a_i} \in \mathcal{A}} \mathbf{a_i}'\mathbf{x}$ for each candidate solution $\mathbf{x}$, which would make the robust formulation considerably more difficult to solve than its nominal counterpart, a linear programming problem. The key insight that preserves the computational tractability of the robust approach is that Problem (4) can be reformulated as a single convex programming problem for any convex uncertainty set $\mathcal{A}$, and specifically, a linear programming problem when $\mathcal{A}$ is a polyhedron (see Ben-Tal and Nemirovski [8]). We now justify this insight by describing the construction of a *tractable, linear* equivalent formulation of Problem (4).

The set $\mathcal{A}$ is defined as follows. To simplify the exposition, we assume that every coefficient $a_{ij}$ of the matrix $\mathbf{A}$ is subject to uncertainty, and that all coefficients are independent. The decision-maker knows range forecasts for all the uncertain parameters, specifically, parameter $a_{ij}$ belongs to a symmetric interval $[\overline{a}_{ij} - \widehat{a}_{ij}, \overline{a}_{ij} + \widehat{a}_{ij}]$ centered at the point forecast $\overline{a}_{ij}$. The half-length $\widehat{a}_{ij}$ measures the precision of the estimate. We define the *scaled deviation* $z_{ij}$ of parameter $a_{ij}$ from its nominal value as:

$$z_{ij} = \frac{a_{ij} - \overline{a}_{ij}}{\widehat{a}_{ij}}. \tag{5}$$

The scaled deviation of a parameter always belongs to $[-1, 1]$.

Although the aggregate scaled deviation for constraint $i$, $\sum_{j=1}^{n} z_{ij}$, could in theory take any value between $-n$ and $n$, the fact that aggregate forecasts are more accurate than individual ones suggests that the "true values" taken by $\sum_{j=1}^{n} z_{ij}$ will belong to a much narrower range. Intuitively, some parameters will exceed their point forecast while others will fall below estimate, so the $z_{ij}$ will tend to cancel each other out. This is illustrated in Figure 1, where we have plotted 50 sample paths of a symmetric random walk over 50 time periods. Figure 1 shows that, when there are few sources of uncertainty (few time periods, little aggregation), the random walk might indeed take its worst-case value; however, as the number of sources of uncertainty increases, this becomes extremely unlikely, as evidenced by the concentration of the sample paths around the mean value of 0. We incorporate

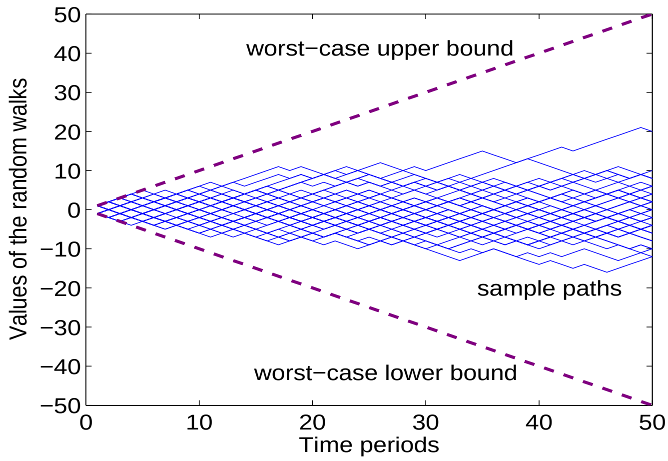


Figure 1: Sample paths as a function of the number of random parameters.

this point in mathematical terms as:

$$\sum_{j=1}^{n} |z_{ij}| \leq \Gamma_i, \ \forall i. \tag{6}$$

The parameter $\Gamma_i$, which belongs to $[0, n]$, is called the *budget of uncertainty* of constraint $i$. If $\Gamma_i$ is integer, it is interpreted as the maximum number of parameters that can deviate from their nominal values.

- If $\Gamma_i = 0$, the $z_{ij}$ for all $j$ are forced to 0, so that the parameters $a_{ij}$ are equal to their point forecasts $\overline{a}_{ij}$ for all $j$ and there is no protection against uncertainty.
- If $\Gamma_i = n$, Constraint (6) is redundant with the fact that $|z_{ij}| \leq 1$ for all $j$. The $i$-th constraint of the problem is completely protected against uncertainty, which yields a very conservative solution.
- If $\Gamma_i \in (0, n)$, the decision-maker makes a trade-off between the protection level of the constraint and the degree of conservatism of the solution.

We provide guidelines to select the budgets of uncertainty at the end of this section. The set $\mathcal{A}$ becomes:

$$\mathcal{A} = \{(a_{ij}) \,|\, a_{ij} = \overline{a}_{ij} + \widehat{a}_{ij}\, z_{ij}, \ \forall i, j, \ \mathbf{z} \in \mathcal{Z}\}\,. \tag{7}$$

with:

$$\mathcal{Z} = \left\{ \mathbf{z} \,\middle|\, |z_{ij}| \leq 1, \ \forall i, j, \ \sum_{j=1}^{n} |z_{ij}| \leq \Gamma_i, \ \forall i \right\}, \tag{8}$$

and Problem (4) can be reformulated as:

$$\begin{aligned} \min \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \overline{\mathbf{a}}_{\mathbf{i}}'\mathbf{x} + \min_{\mathbf{z_i} \in \mathcal{Z}_i} \sum_{j=1}^{n} \widehat{a}_{ij}\, x_j\, z_{ij} \geq b_i, \quad \forall i, \\ & \mathbf{x} \in X, \end{aligned} \tag{9}$$

where $\mathbf{z_i}$ is the vector whose $j$-th element is $z_{ij}$ and $\mathcal{Z}_i$ is defined as:

$$\mathcal{Z}_i = \left\{ \mathbf{z_i} \,\middle|\, |z_{ij}| \leq 1,\ \forall j,\ \sum_{j=1}^{n} |z_{ij}| \leq \Gamma_i \right\}. \tag{10}$$

$\min_{\mathbf{z_i} \in \mathcal{Z}_i} \sum_{j=1}^{n} \widehat{a}_{ij}\, x_j\, z_{ij}$ for a given $i$ is equivalent to:

$$\begin{aligned} -\max \quad & \sum_{j=1}^{n} \widehat{a}_{ij}\, |x_j|\, z_{ij} \\ \text{s.t.} \quad & \sum_{j=1}^{n} z_{ij} \leq \Gamma_i, \\ & 0 \leq z_{ij} \leq 1, \qquad \forall j, \end{aligned} \tag{11}$$

which is linear in the decision vector $\mathbf{z_i}$. Applying strong duality arguments to Problem (11) (see Bertsimas and Sim [18] for details), we then reformulate the robust problem as a **linear programming problem**:

$$\begin{aligned} \min \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \overline{\mathbf{a}}_{\mathbf{i}}'\mathbf{x} - \Gamma_i\, p_i - \sum_{j=1}^{n} q_{ij} \geq b_i, && \forall i, \\ & p_i + q_{ij} \geq \widehat{a}_{ij}\, y_j, && \forall i, j, \\ & -y_j \leq x_j \leq y_j, && \forall j, \\ & p_i,\ q_{ij} \geq 0, && \forall i, j, \\ & \mathbf{x} \in X, \end{aligned} \tag{12}$$

With $m$ the number of constraints subject to uncertainty and $n$ the number of variables in the deterministic problem (1), Problem (12) has $n+m(n+1)$ new variables and $n(m+2)$ new constraints besides nonnegativity. An appealing feature of this formulation is that linear programming problems can be solved efficiently, including by the commercial software used in industry.

At optimality,

1. $y_j$ will equal $|x_j|$ for any $j$,

2. $p_i$ will equal the $\lceil \Gamma_i \rceil$-th greatest $\widehat{a}_{ij}\, |x_j|$, for any $i$,

3. $q_{ij}$ will equal $\widehat{a}_{ij}\,|x_j| - p_i$ if $\widehat{a}_{ij}\,|x_j|$ is among the $\lfloor\Gamma_i\rfloor$-th greatest $\widehat{a}_{ik}\,|x_k|$ and 0 otherwise, for any $i$ and $j$. (Equivalently, $q_{ij} = \max(0, \widehat{a}_{ij}\,|x_j| - p_i)$.)

To implement this framework, the decision-maker must now assign a value to the budget of uncertainty $\Gamma_i$ for each $i$. The values of the budgets can for instance reflect the manager's own attitude towards uncertainty; the connection between risk preferences and uncertainty sets is studied in depth in Section 4. Here, we focus on selecting the budgets so that the constraints $\mathbf{Ax} \geq \mathbf{b}$ are satisfied with high probability in practice, despite the lack of precise information on the distribution of the random matrix $\mathbf{A}$. The central result linking the value of the budget to the probability of constraint violation is due to Bertsimas and Sim [18] and can be summarized as follows:

For the constraint $\mathbf{a_i'x} \geq b_i$ to be violated with probability at most $\epsilon_i$, when each $a_{ij}$ obeys a symmetric distribution centered at $\overline{a}_{ij}$ and of support $[\overline{a}_{ij} - \widehat{a}_{ij}, \overline{a}_{ij} + \widehat{a}_{ij}]$, it is sufficient to choose $\Gamma_i$ at least equal to $1 + \Phi^{-1}(1-\epsilon_i)\,\sqrt{n}$, where $\Phi$ is the cumulative distribution of the standard Gaussian random variable.

As an example, for $n = 100$ sources of uncertainty and $\epsilon_i = 0.05$ in constraint $i$, $\Gamma_i$ must be at least equal to 17.4, i.e., it is sufficient to protect the system against only 18% of the uncertain parameters taking their worst-case value. Most importantly, $\Gamma_i$ is always of the order of $\sqrt{n}$. Therefore, the constraint can be protected with high probability while keeping the budget of uncertainty, and hence the degree of conservatism of the solution, moderate.

We now illustrate the approach on a few simple examples.

**Example 2.1 (Portfolio management, Bertsimas and Sim [18])** *A decision-maker must allocate his wealth among* 150 *assets in order to maximize his return. He has established that the return of asset* $i$ *belongs to the interval* $[r_i - s_i, r_i + s_i]$ *with* $r_i = 1.15 + i\,(0.05/150)$ *and* $s_i = (0.05/450)\sqrt{300 \cdot 151 \cdot i}$. *Short sales are not allowed. Obviously, in the deterministic problem where all returns are equal to their point forecasts, it is optimal to invest everything in the asset with the greatest nominal return, here, asset* 150. *(Similarly, in the conservative approach where all returns equal their worst-case values, it is optimal to invest everything in the asset with the greatest worst-case return, which is asset* 1.*)*

*Figure 2 depicts the minimum budget of uncertainty required to guarantee an appropriate performance for the investor, in this context meaning that the actual value of his portfolio will exceed the value predicted by the robust optimization model with probability at least equal to the numbers on the x-axis. We note that performance requirements of up to 98% can be achieved by a small budget of uncertainty (*$\Gamma \approx 26$*, protecting about 17% of the sources of randomness), but more stringent constraints require a drastic increase in the protection level, as evidenced by the almost vertical increase* *in the curve. The investor would like to find a portfolio allocation such that there is only a probability*

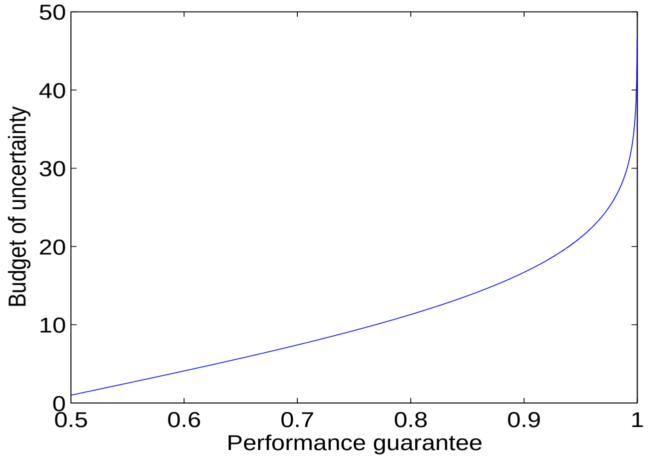


Figure 2: Minimum budget of uncertainty to ensure performance guarantee.

*of 5% that the actual portfolio value will fall below the value predicted by his optimization model. Therefore, he picks* $\Gamma \geq 21.15$*, e.g.,* $\Gamma = 22$*, and solves the linear programming problem:*

$$
\begin{aligned}
\max \quad & \sum_{i=1}^{150} r_i\, x_i - \Gamma p - \sum_{i=1}^{150} q_i \\
s.t. \quad & \sum_{i=1}^{150} x_i = 1, \\
& p + q_i \geq s_i\, x_i, \ \ \forall i, \\
& p,\, q_i,\, x_i \geq 0, \ \ \forall i.
\end{aligned}
\tag{13}
$$

*At optimality, he invests in every asset, and the fraction of wealth invested in asset* $i$ *decreases from 4.33% to 0.36% as the index* $i$ *increases from 1 to 150. The optimal objective is 1.1452.*

*To illustrate the impact of the robust methodology, assume the true distribution of the return of asset* $i$ *is Gaussian with mean* $r_i$ *and standard deviation* $s_i/2$*, so that the range forecast for return* $i$ *includes every value within two standard deviations of the mean. Asset returns are assumed to be independent. Then:*

- *The portfolio value in the* nominal *strategy, where everything is invested in asset* 150*, obeys a Gaussian distribution with mean 1.2 and standard deviation 0.1448.*

- *The portfolio value in the* conservative *strategy, where everything is invested in asset* 1*, obeys a Gaussian distribution with mean 1.1503 and standard deviation 0.0118.*

- *The portfolio value in the* robust *strategy, which leads to a diversification of the investor's holdings, obeys a Gaussian distribution with mean 1.1678 and standard deviation 0.0063.*

*Hence, not taking uncertainty into account rather than implementing the robust strategy increases risk (measured by the standard deviation) by a factor of 23 while yielding an increase in expected return of only 2.7%, and being too pessimistic regarding the outcomes doubles the risk and also decreases the expected return.*

**Example 2.2 (Inventory management, Thiele [45])** *A warehouse manager must decide how many products to order, given that the warehouse supplies $n$ stores and it is only possible to order once for the whole planning period. The warehouse has an initial inventory of zero, and incurs a unit shortage cost $s$ per unfilled item and a unit holding cost $h$ per item remaining in the warehouse at the end of the period. Store demands are assumed to be i.i.d. with a symmetric distribution around the mean and all of the stores have the same range forecast $[\overline{w} - \widehat{w}, \overline{w} + \widehat{w}]$ with $\overline{w}$ the nominal forecast, common to each store. Let $x$ be the number of items ordered by the decision-maker, whose goal is to minimize the total cost $\max\{h(x - \sum_{i=1}^n w_i), s(\sum_{i=1}^n w_i - x)\}$, with $\sum_{i=1}^n w_i$ the actual aggregate demand. The robust problem for a given budget of uncertainty $\Gamma$ can be formulated as:*

$$\begin{aligned} \min \quad & Z \\ s.t. \quad & Z \geq h(x - n\overline{w} + \Gamma\widehat{w}), \\ & Z \geq s(-x + n\overline{w} + \Gamma\widehat{w}), \\ & x \geq 0. \end{aligned} \tag{14}$$

*The solution to Problem (14) is available in closed form and is equal to:*

$$x_\Gamma = n\overline{w} + \frac{s-h}{s+h}\Gamma\widehat{w}. \tag{15}$$

*The optimal objective is then:*

$$C_\Gamma = \frac{2\,h\,s}{s+h}\Gamma\widehat{w}. \tag{16}$$

*If shortage is more penalized than holding, the decision-maker will order more than the nominal aggregate forecast, and the excess amount will be proportional to the maximum deviation $\Gamma\,\widehat{w}$, as well as the ratio $(s-h)/(s+h)$. The optimal order is linear in the budget of uncertainty.*

*Using the Central Limit Theorem, and assuming that the variance of each store demand is known and equal to $\sigma^2$, it is straightforward to show that the optimal objective $C_\Gamma$ is an upper bound to the true cost with probability $1-\epsilon$ when $\Gamma$ is at least equal to $(\sigma/\widehat{w})\sqrt{n}\,\Phi^{-1}(1-\epsilon/2)$. This formula is* *independent of the cost parameters h and s. For instance with* $n = 100$ *and* $\widehat{w} = 2\,\sigma$*, the actual cost falls below* $C_{10}$ *with probability* $0.95$*.*

*Because in this case the optimal solution is available in closed form, we can analyze in more depth the impact of the budget of uncertainty on the practical performance of the robust solution. To illustrate the two dangers of "not worrying enough" about uncertainty (i.e., only considering the nominal values of the parameters) and "worrying too much" (i.e., only considering their worst-case values) in practical implementations, we compute the expected cost for the worst-case probability distribution of the aggregate demand* $W$*. We only use the following information on* $W$*: its distribution is symmetric with mean* $n\,\overline{w}$ *and support* $[n(\overline{w}-\widehat{w}), n(\overline{w}+\widehat{w})]$*, and (as established by Bertsimas and Sim [18])* $W$ *falls within* $[n\overline{w} - \Gamma\widehat{w}, n\overline{w} + \Gamma\widehat{w}]$ *with probability* $2\,\phi - 1$ *where* $\phi = \Phi((\Gamma - 1)/\sqrt{n})$*. Let* $\mathcal{W}$ *be the set of probability distributions satisfying these assumptions. Thiele [45] proves the following bound:*

$$\max_{W \in \mathcal{W}} E[\max\{h(x-W), s(W-x)\}] = \widehat{w}(s+h)\left[n(1-\phi) + \Gamma\left\{\phi - \frac{s^2+h^2}{(s+h)^2}\right\}\right]. \quad (17)$$

*In Figure 3, we plot this upper bound on the expected cost for* $n = 100$*,* $\widehat{w} = 1$*,* $h = 1$ *and* $s = 2$*, 3 and 4. We note that not incorporating uncertainty in the model is the more costly mistake the*

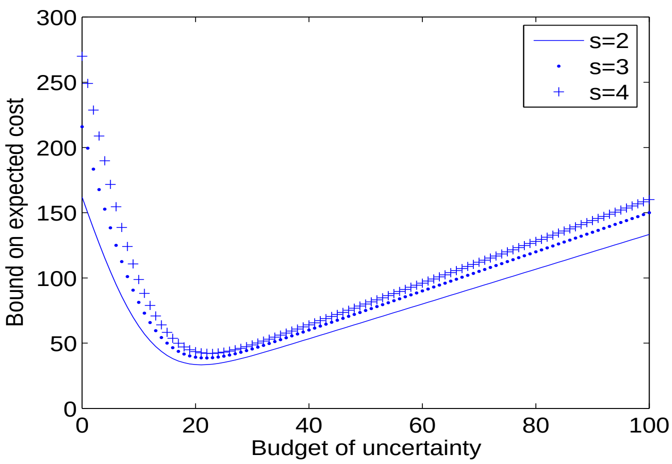


Figure 3: Maximum expected cost as a function of the budget of uncertainty.

*manager can make in this setting (as opposed to being too conservative), the penalty increases when the shortage cost increases. The budget of uncertainty minimizing this bound is approximately equal to 20 and does not appear to be sensitive to the value of the cost parameters.*

*The key insight of Figure 3 is that accounting for a* limited *amount of uncertainty via the robust optimization framework leads to significant cost benefits. A decision-maker implementing the nominal* *strategy will be penalized for not planning at all for randomness, i.e., the aggregate demand deviating from its point forecast, but protecting the system against the most negative outcome will also result in lost profit opportunities. The robust optimization approach achieves a trade-off between these two extremes.*

## 2.2 Extensions

### 2.2.1 Discrete decision variables

The modeling power of robust optimization also extends to discrete decision variables. Integer decision variables can be incorporated into the set $X$ (which is then no longer a polyhedron), while binary variables allow for the development of a specifically tailored algorithm due to Bertsimas and Sim [17]. We describe this approach for the binary programming problem:

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}'\mathbf{x} \leq b \\ & \mathbf{x} \in \{0,1\}^n. \end{aligned} \tag{18}$$

Problem (18) can be interpreted as a *capital allocation problem* where the decision-maker must choose between $n$ projects to maximize his payoff under a budget constraint, but does not know exactly how much money each project will require. In this setting, the robust problem (12) (modified to take into account the sign of the inequality and the maximization) becomes:

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \overline{\mathbf{a}}'\mathbf{x} + \Gamma\, p + \sum_{j=1}^{n} q_j \leq b \\ & p + q_j \geq \widehat{a}_j\, x_j, \ \forall j, \\ & p \geq 0,\ \mathbf{q} \geq \mathbf{0}, \\ & \mathbf{x} \in \{0,1\}^n. \end{aligned} \tag{19}$$

As noted for Problem (12), at optimality $q_j$ will equal $\max(0, \widehat{a}_j\, x_j - p)$. The major insight here is that, since $x_j$ is binary, $q_j$ can take only two values: $\max(0, \widehat{a}_j - p)$ and 0, which can be rewritten as $\max(0, \widehat{a}_j - p)\, x_j$. Therefore, the optimal $p$ will be one of the $\widehat{a}_j$ and the optimal solution can be found by solving $n$ subproblems *of the same size and structure* as the original deterministic problem, and keeping the one with the highest objective. Solving these subproblems can be automated with no difficulty, for instance in AMPL/CPLEX, thus preserving the computational tractability of the robust optimization approach. Subproblem $i$, $i = 1, \ldots, n$, is defined as the following **binary**

**programming problem**:

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \overline{\mathbf{a}}'\mathbf{x} + \sum_{j=1}^{n} \max(0, \widehat{a}_j - \widehat{a}_i)\, x_j \leq b - \Gamma\, \widehat{a}_i \\ & \mathbf{x} \in \{0,1\}^n. \end{aligned} \tag{20}$$

It has the same number of constraints and decision variables as the original problem.

**Example 2.3 (Capital allocation, Bertsimas and Sim [17])** *The manager has a budget* $b$ *of* \$4,000 *and can choose between 200 projects. The nominal amount of money* $\overline{a}_i$ *required to complete project* $i$ *is chosen randomly from the set* $\{20, \ldots, 29\}$*, the range forecast allows for a deviation of at most 10% of this estimate. The value (or importance)* $c_i$ *of project* $i$ *is chosen randomly from* $\{16, \ldots, 77\}$*. Bertsimas and Sim [17] show that, while the nominal problem yields an optimal value of* 5,592*, taking* $\Gamma$ *equal to* 37 *ensures that the decision-maker will remain within budget with a probability of* 0.995*, and with a decrease in the objective value of only 1.5%. Therefore, the system can be protected against uncertainty at very little cost.*

### 2.2.2 Generic polyhedral uncertainty sets and norms

Since the main mathematical tool used in deriving tractable robust formulations is the use of strong duality in linear programming, it should not be surprising that the robust counterparts to linear problems with generic polyhedral uncertainty sets remain linear. For instance, if the set $\mathcal{Z}_i$ for constraint $i$ is defined by: $\mathcal{Z}_i = \{\mathbf{z} \,|\, \mathbf{F_i}|\mathbf{z}| \leq \mathbf{g_i},\ |\mathbf{z}| \leq \mathbf{e}\}$ where $\mathbf{e}$ is the unit vector, rather than $\mathcal{Z}_i = \{\mathbf{z} \,|\, \sum_{j=1}^{n_i} |z_{ij}| \leq \Gamma_i,\ |z_{ij}| \leq 1,\ \forall j\}$, it is immediately possible to formulate the robust problem as:

$$\begin{aligned} \min \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \overline{\mathbf{a}}_\mathbf{i}'\,\mathbf{x} - \mathbf{g}_\mathbf{i}'\,\mathbf{p_i} - \mathbf{e}'\mathbf{q_i} \geq b_i, \quad \forall i, \\ & \mathbf{F}_\mathbf{i}'\,\mathbf{p_i} + \mathbf{q_i} \geq (\mathbf{diag}\,\widehat{\mathbf{a}}_\mathbf{i})\,\mathbf{y}, \quad \forall i, \\ & -\mathbf{y} \leq \mathbf{x} \leq \mathbf{y}, \\ & \mathbf{p},\, \mathbf{q} \geq \mathbf{0}, \\ & \mathbf{x} \in X, \end{aligned} \tag{21}$$

Moreover, given that the precision of each individual forecast $\overline{a}_{ij}$ is quantified by the parameter $\widehat{a}_{ij}$, which measures the maximum "distance" of the true scalar parameter $a_{ij}$ from its nominal value $\overline{a}_{ij}$, it is natural to take this analysis one step further and consider the distance of the true vector of parameters $\mathbf{A}$ from its point forecast $\overline{\mathbf{A}}$. Uncertainty sets arising from limitations on the distance (measured by an arbitrary norm) between uncertain coefficients and their nominal values have been investigated by Bertsimas et. al. [16], who show that reframing the uncertainty set in those terms lead to convex problems with constraints involving a dual norm, and provide a unified treatment of robust optimization as described by Ben-Tal and Nemirovski [7], [8], El-Ghaoui et. al. [27], [28] and Bertsimas and Sim [18]. Intuitively, robust optimization protects the system against any value of the parameter vector within a prespecified "distance" from its point forecast.

### 2.2.3 Additional models and applications

Robust optimization has been at the center of many research efforts over the last decade, and in this last paragraph we mention a few of those pertaining to static decision-making under uncertainty for the interested reader. This is, of course, far from an exhaustive list.

While this tutorial focuses on linear programming and polyhedral uncertainty sets, the robust optimization paradigm is well-suited to a much broader range of problems. Atamturk [2] provides strong formulations for robust mixed 0-1 programming under uncertainty in the objective coefficients. Sim [42] extends the robust framework to quadratically constrained quadratic problems, conic problems as well as semidefinite problems, and provides performance guarantees. Ben-Tal et. al. [10] consider tractable approximations to robust conic-quadratic problems. An important application area is portfolio management, where Goldfarb and Iyengar [29] protect the optimal asset allocation from estimation errors in the parameters by using robust optimization techniques. Ordonez and Zhao [34] apply the robust framework to the problem of expanding network capacity when demand and travel times are uncertain. Finally, Ben-Tal et. al. [4] investigate robust problems where the decision-maker requires a controlled deterioration of the performance when the data falls outside the uncertainty set.

# 3 Dynamic Decision-Making under Uncertainty

## 3.1 Generalities

Section 2 has established the power of robust optimization in static decision-making, where it immunizes the solution against infeasibility and suboptimality. We now extend our presentation to the dynamic case. In this setting, information is revealed *sequentially* over time, and the manager makes a *series of decisions*, which take into account the historical realizations of the random variables. Because dynamic optimization involves multiple decision epochs and must capture the wide range of circumstances (i.e., state of the system, values taken by past sources of randomness) in which decisions are made, the fundamental issue here is one of *computational tractability*.

Multi-stage stochastic models provide an elegant theoretical framework to incorporate uncertainty revealed over time (see, e.g., Bertsekas [11] for an introduction.) However, the resulting large-scale formulations quickly become intractable as the size of the problem increases, thus limiting the practical usefulness of these techniques. For instance, a manager planning for the next quarter (13 weeks) and considering 3 values of the demand each week (high, low or medium) has just created $3^{13} \approx 1.6$ million scenarios in the stochastic framework. Approximation schemes such as neuro-dynamic programming (Bertsekas and Tsitsiklis [12]) have yet to be widely implemented, in part because of the difficulty in finetuning the approximation parameters. Moreover, as in the static case, each scenario needs to be assigned a specific probability of occurrence, and the difficulty in estimating these parameters accurately is compounded in multi-stage problems by long time horizons. Intuitively, "one can predict tomorrow's value of the Dow Jones Industrial Average more accurately than next year's value." (Nahmias [32])

Therefore, a decision-maker using a stochastic approach might expand considerable computational resources to solve a multi-stage problem, which will *not* be the true problem he is confronted with because of estimation errors. A number of researchers have attempted to address this issue by implementing robust techniques directly in the stochastic framework (i.e., optimizing over the worst-case probabilities in a set), e.g., Zackova-Dupacova [48], [26], Shapiro [40] for two-stage stochastic programming and Iyengar [30], Nilim and El-Ghaoui [33] for multi-stage dynamic programming. Although this method protects the system against parameter ambiguity, it suffers from the same limitations as the algorithm with perfect information; hence, if a problem relying on a probabilistic description of the uncertainty is computationally intractable, its robust counterpart will be intractable as well.

In contrast, we approach dynamic optimization problems subject to uncertainty by representing the *random variables*, rather than the underlying probabilities, as uncertain parameters belonging to given uncertainty sets. This is in line with the methodology presented in the static case. The extension of the approach to dynamic environments raises the following questions:

1. Is the robust optimization paradigm *tractable* in dynamic settings?
2. Does the manager derive *deeper insights* into the impact of uncertainty?
3. Can the methodology incorporate the *additional information* received by the decision-maker over time?

As explained below, the answer to each of these three questions is *yes*.

## 3.2 A First Model

A first, intuitive approach is to incorporate uncertainty to the underlying *deterministic* formulation. In this tutorial, we focus on applications that can be modeled (or approximated) as linear programming problems when there is no randomness. For clarity, we present the framework in the context of inventory management; the exposition closely follows Bertsimas and Thiele [20].

### 3.2.1 Scalar case

We start with the simple case where the decision-maker must decide how many items to order at each time period at a single store. (In mathematical terms, the state of the system can be described as a scalar variable, specifically, the amount of inventory in the store.) We use the following notation:

$x_t$: inventory at the beginning of time period $t$,

$u_t$: amount ordered at the beginning of time period $t$,

$w_t$: demand occurring during time period $t$.

Demand is backlogged over time, and orders made at the beginning of a time period arrive at the end of that same period. Therefore, the dynamics of the system can be described as a *linear* equation:

$$x_{t+1} = x_t + u_t - w_t, \tag{22}$$

which yields the closed-form formula:

$$x_{t+1} = x_0 + \sum_{\tau=0}^{t} (u_\tau - w_\tau). \tag{23}$$

The cost incurred at each time period has two components:

1. An ordering cost linear in the amount ordered, with $c$ the unit ordering cost (Bertsimas and Thiele [20] also consider the case of a fixed cost charged whenever an order is made),

2. An inventory cost, with $h$, respectively $s$, the unit cost charged per item held in inventory, resp. backlogged, at the end of each time period.

The decision-maker seeks to minimize the total cost over a time horizon of length $T$. He has a range forecast $[\overline{w}_t - \widehat{w}_t, \overline{w}_t + \widehat{w}_t]$, centered at the nominal forecast $\overline{w}_t$, for the demand at each time period $t$, with $t = 0, \ldots, T-1$. If there is no uncertainty, the problem faced by the decision-maker can be formulated as a linear programming problem:

$$
\begin{aligned}
\min \quad & c \sum_{t=0}^{T-1} u_t + \sum_{t=0}^{T-1} y_t \\
\text{s.t.} \quad & y_t \geq h \left( x_0 + \sum_{\tau=0}^{t} (u_\tau - \overline{w}_\tau) \right), \qquad \forall t \\
& y_t \geq -s \left( x_0 + \sum_{\tau=0}^{t} (u_\tau - \overline{w}_\tau) \right), \quad \forall t, \\
& u_t \geq 0, \qquad\qquad\qquad\qquad\qquad \forall t.
\end{aligned}
\tag{24}
$$

At optimality, $y_t$ is equal to the inventory cost computed at the end of time period $t$, i.e., $\max(h\, x_{t+1}, -s\, x_{t+1})$. The optimal solution to Problem (24) is to order nothing if there is enough in inventory at the beginning of period $t$ to meet the demand $\overline{w}_t$, and order the missing items, i.e., $\overline{w}_t - x_t$, otherwise, which is known in inventory management as a *(S,S) policy* with basestock level $\overline{w}_t$ at time $t$. (The basestock level quantifies the amount of inventory on hand or on order at a given time period, see, e.g., Porteus [35].)

The robust optimization approach consists in replacing each deterministic demand $\overline{w}_t$ by an uncertain parameter $w_t = \overline{w}_t + \widehat{w}_t\, z_t$, $|z_t| \leq 1$, for all $t$, and guaranteeing that the constraints hold for any scaled deviations belonging to a given uncertainty set. Because the constraints depend on the time period, the uncertainty set will depend on the time period as well, and specifically, the amount of uncertainty faced by the cumulative demand up to (and including) time $t$. This motivates introducing a *sequence* of budgets of uncertainty $\Gamma_t$, $t = 0, \ldots, T-1$, rather than using a single budget as in the static case. Natural requirements for such a sequence are that the budgets increase over time, as uncertainty increases with the length of the time horizon considered, and do not increase by more than one at each time period, since only one new source of uncertainty is revealed at any time.

Let $\overline{x}_t$ be the amount in inventory at time $t$ if there is no uncertainty: $\overline{x}_{t+1} = x_0 + \sum_{\tau=0}^{t}(u_\tau - \overline{w}_\tau)$ for all $t$. Also, let $Z_t$ be the optimal solution of:

$$
\begin{aligned}
\max \quad & \sum_{\tau=0}^{t} \widehat{w}_\tau\, z_\tau \\
\text{s.t.} \quad & \sum_{\tau=0}^{t} z_\tau \leq \Gamma_t, \\
& 0 \leq z_\tau \leq 1, \ \forall \tau \leq t.
\end{aligned}
\tag{25}
$$

From $0 \leq \Gamma_t - \Gamma_{t-1} \leq 1$, it is straightforward to show that $0 \leq Z_t - Z_{t-1} \leq \widehat{w}_t$ for all $t$. The robust counterpart to Problem (24) can be formulated as a **linear programming problem:**

$$
\begin{aligned}
\min \quad & \sum_{t=0}^{T-1} (c\, u_t + y_t) \\
\text{s.t.} \quad & y_t \geq h(\overline{x}_{t+1} + Z_t), \qquad \forall t \\
& y_t \geq s(-\overline{x}_{t+1} + Z_t), \quad \forall t, \\
& \overline{x}_{t+1} = \overline{x}_t + u_t - \overline{w}_t, \quad \forall t, \\
& u_t \geq 0, \qquad \forall t.
\end{aligned}
\tag{26}
$$

A key insight in the analysis of the robust optimization approach is that Problem (26) is equivalent to a deterministic inventory problem where the demand at time $t$ is defined by:

$$
w'_t = \overline{w}_t + \frac{s-h}{s+h}(Z_t - Z_{t-1}). \tag{27}
$$

Therefore, the optimal robust policy is $(\mathbf{S}, \mathbf{S})$ **with basestock level $\mathbf{w'_t}$.** We make the following observations on the robust basestock levels:

- They do not depend on the unit ordering cost, and depend on the holding and shortage costs only through the ratio $(s-h)/(s+h)$.
- They remain higher, respectively lower, than the nominal ones over the time horizon when shortage is penalized more, respectively less, than holding, and converge towards their nominal values as the time horizon increases.
- They are not constant over time, even when the nominal demands are constant, because they also capture information on the time elapsed since the beginning of the planning horizon.
- They are closer to the nominal basestock values than those obtained in the robust myopic approach (when the robust optimization model only incorporates the next time period); hence, taking into account the whole time horizon mitigates the impact of uncertainty at each time period.

Bertsimas and Thiele [20] provide guidelines to select the budgets of uncertainty based on the worst-case expected cost computed over the set of random demands with given mean and variance. For instance, when $c = 0$ (or $c \ll h$, $c \ll s$), and the random demands are i.i.d. with mean $\overline{w}$ and standard deviation $\sigma$, they take:

$$\Gamma_t = \min\left(\frac{\sigma}{\widehat{w}}\sqrt{\frac{t+1}{1-\alpha^2}}, t+1\right), \tag{28}$$

with $\alpha = (s-h)/(s+h)$. Equation (28) suggests two phases in the decision-making process:

1. an early phase where the decision-maker takes a very conservative approach ($\Gamma_t = t+1$),

2. a later phase where the decision-maker takes advantage of the aggregation of the sources of randomness ($\Gamma_t$ proportional to $\sqrt{t+1}$).

This is in line with the empirical behavior of the uncertainty observed in Figure 1.

**Example 3.1 (Inventory management, Bertsimas and Thiele [20])**
*For i.i.d. demands with mean* 100*, standard deviation* 20*, range forecast* $[60, 140]$*, a time horizon of 20 periods and cost parameters* $c = 0$*,* $h = 1$*,* $s = 3$*, the optimal basestock level is given by:*

$$w_t' = 100 + \frac{20}{\sqrt{3}}(\sqrt{t+1} - \sqrt{t}), \tag{29}$$

*which decreases approximately as* $1/\sqrt{t}$*. Here, the basestock level decreases from 111.5 (for* $t = 0$*) to 104.8 (for* $t = 2$*) to 103.7 (for* $t = 3$*), and ultimately reaches* 101.3 *(*$t = 19$*.)*

The robust optimization framework can incorporate a wide range of additional features, including fixed ordering costs, fixed lead times, integer order amounts, capacity on the orders and capacity on the amount in inventory.

### 3.2.2 Vector case

We now extend the approach to the case where the decision-maker manages multiple components of the supply chain, such as warehouses and distribution centers. In mathematical terms, the state of the system is described by a vector. While traditional stochastic methods quickly run into tractability issues when the dynamic programming equations are multi-dimensional, we will see that the robust optimization framework incorporates randomness with no difficulty, in the sense that it can be solved as efficiently as its nominal counterpart. In particular, the robust counterpart of the deterministic inventory management problem remains a linear programming problem, for any topology of the underlying supply network.

We first consider the case where the system is faced by only one source of uncertainty at each time period, but the state of the system is now described by a vector. A classical example in inventory management arises in *series systems*, where goods proceed through a number of stages (factory, distributor, wholesaler, retailer) before being sold to the customer. We define stage $k$, $k = 1, \ldots, N$, as the stage in which the goods are $k$ steps away from exiting the network, with stage $k+1$ supplying stage $k$ for $1 \leq k \leq N-1$. Stage 1 is the stage subject to customer demand uncertainty and stage $N$ has an infinite supply of goods. Stage $k$, $k \leq N-1$, cannot supply to the next stage more items that it currently has in inventory, which introduces coupling constraints between echelons in the mathematical model. In line with Clark and Scarf [24], we compute the inventory costs at the *echelon* level, with echelon $k$, $1 \leq k \leq N$, being defined as the union of all stages from 1 to $k$ as well as the links in-between. For instance, when the series system represents a manufacturing line where raw materials become work-in-process inventory and ultimately finished products, holding and shortage costs are incurred for items that have reached and possibly moved beyond a given stage in the manufacturing process. Each echelon has the same structure as the single stage described in Section 3.2.1, with echelon-specific cost parameters.

Bertsimas and Thiele [20] show that:

1. The robust optimization problem can be reformulated as a linear programming problem when there are no fixed ordering costs and a mixed-integer programming problem otherwise.

2. The optimal policy for echelon $k$ in the robust problem is the same as in a deterministic single-stage problem with modified demand at time $t$:

$$w'_t = \overline{w}_t + \frac{p_k - h_k}{p_k + h_k}(Z_t - Z_{t-1}), \tag{30}$$

   with $Z_t$ defined as in Equation (25), and *time-varying capacity* on the orders.

3. When there is no fixed ordering cost, the optimal policy for echelon $k$ is the same as in a deterministic *uncapacitated* single-stage problem with demand $w'_t$ at time $t$ and *time-varying cost coefficients*, which depend on the Lagrange multipliers of the coupling constraints. In particular, the policy is basestock.

Hence, the robust optimization approach provides theoretical insights into the impact of uncertainty on the series system, and recovers the optimality of basestock policies established by Clark and Scarf [24] in the stochastic programming framework when there is no fixed ordering costs. It also allows the decision-maker to incorporate uncertainty and gain a deeper understanding of problems for which the optimal solution in the stochastic programming framework is *not known*, such as more complex hierarchical networks. Systems of particular interest are those with an expanding tree structure, as the decision-maker can still define echelons in this context and derive some properties on the structure of the optimal solution. Bertsimas and Thiele [20] show that the insights gained for series systems extend to tree networks, where the demand at the retailer is replaced by the cumulative demand at that time period for all retailers in the echelon.

**Example 3.2 (Inventory management, Bertsimas and Thiele [20])** *A decision-maker implements the robust optimization approach on a simple tree network with one warehouse supplying two stores. Ordering costs are all equal to 1, holding and shortage costs at the stores are all equal to 8, while the holding, respectively shortage, costs for the whole system is 5, respectively 7. Demands at the store are i.i.d. with mean 100, standard deviation 20 and range forecast [60,140]. The stores differ by their initial inventory: 150 and 50 items, respectively, while the whole system initially has 300 items. There are 5 time periods. Bertsimas and Thiele [20] compare the sample cost of the robust approach with a myopic policy, which adopts a probabilistic description of the randomness at the expense of the time horizon. Figure 4 shows the costs when the myopic policy assumes Gaussian distributions at both stores, which in reality are Gamma with the same mean and variance. Note that the graph for the robust policy is shifted to the left (lower costs) and is narrower than the one for the myopic approach (less volatility).*

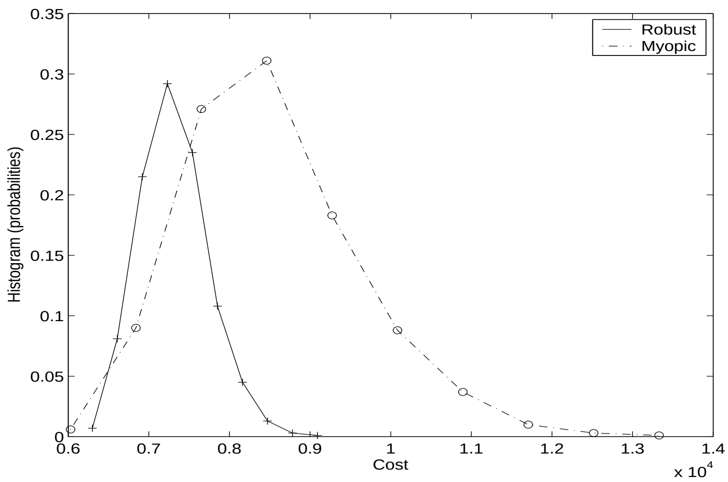


Figure 4: Comparison of costs of robust and myopic policy.

*While the error in estimating the distributions to implement the myopic policy is rather small, Figure 4 indicates that not taking into account the time horizon significantly penalizes the decision-maker, even for short horizons as in this example. Figure 5 provides more insights into the impact of the time horizon on the optimal costs. In particular, the distribution of the relative performance between robust and myopic policies shifts to the right of the threshold 0 and becomes narrower (consistently better performance for the robust policy) as the time horizon increases.*

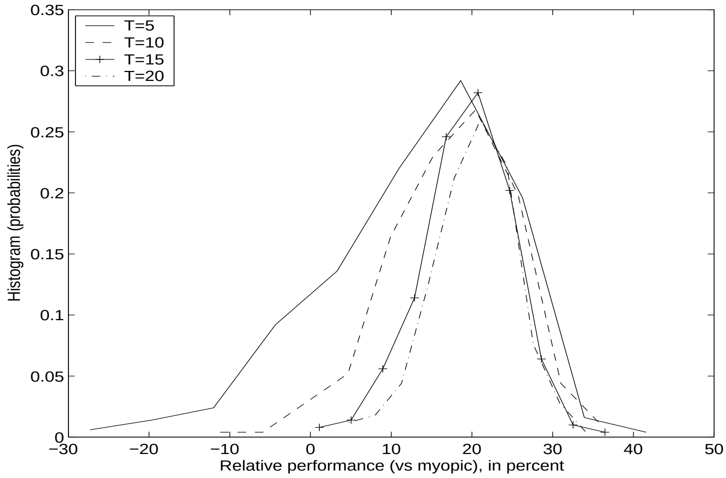


Figure 5: Impact of the time horizon.

*These results suggest that taking randomness into account* throughout the time horizon *plays a more important role on system performance than having a detailed probabilistic knowledge of the uncertainty for the next time period.*

### 3.2.3 Dynamic budgets of uncertainty

In general, the robust optimization approach we have proposed in Section 3.2 does not naturally yield policies in dynamic environments and must be implemented on a rolling horizon basis, i.e., the robust problem must be solved repeatedly over time to incorporate new information. In this section, we introduce an extension of this framework proposed by Thiele [46], which (i) allows the decision-maker to obtain *policies*, (ii) emphasizes the connection with Bellman's *recursive equations* in stochastic dynamic programming, and (iii) identifies the sources of randomness that affect the system *most negatively.* We present the approach when both state and control variables are scalar and there is only one source of uncertainty at each time period. With similar notation as in Section 3.2.2, the state variable obeys the linear dynamics given by:

$$x_{t+1} = x_t + u_t - w_t, \ \forall t = 0, \dots, T-1. \tag{31}$$

The set of allowable control variables at time $t$ for any state $x_t$ is defined as $U_t(x_t)$. The random variable $w_t$ is modeled as an uncertain parameter with range forecast $[\overline{w}_t - \widehat{w}_t, \overline{w}_t + \widehat{w}_t]$; the decision-maker seeks to protect the system against $\Gamma$ sources of uncertainty taking their worst-case value over the time horizon. The cost incurred at each time period is the sum of state costs $f_t(x_t)$ and control costs $g_t(u_t)$, where both functions $f_t$ and $g_t$ are convex for all $t$. Here, we assume that the state costs are computed at the beginning of each time period for simplicity.

The approach hinges on the following question: how should the decision-maker spend a budget of uncertainty of $\Gamma$ units given to him at time 0, and specifically, for any time period, should he spend one unit of his remaining budget to protect the system against the present uncertainty or keep all of it for future use? In order to identify the time periods (and states) the decision-maker should use his budget on, we consider only three possible values for the uncertain parameter at time $t$: nominal, highest and smallest. Equivalently, $w_t = \overline{w}_t + \widehat{w}_t\, z_t$ with $z_t \in \{-1, 0, 1\}$. The **robust counterpart to Bellman's recursive equations** for $t \leq T-1$ is then defined as:

$$J_t(x_t, \Gamma_t) = f_t(x_t) + \min_{u_t \in U_t(x_t)} \left[ g_t(u_t) + \max_{z_t \in \{-1,0,1\}} J_t(\overline{x}_{t+1} - \widehat{w}_t\, z_t, \Gamma_t - |z_t|) \right], \; \Gamma_t \geq 1, \quad (32)$$

$$J_t(x_t, 0) = f_t(x_t) + \min_{u_t \in U_t(x_t)} \left[ g_t(u_t) + J_t(\overline{x}_{t+1}, 0) \right]. \quad (33)$$

with the notation $\overline{x}_{t+1} = x_t + u_t - \overline{w}_t$, i.e., $\overline{x}_{t+1}$ is the value taken by the state at the next time period if there is no uncertainty. We also have the boundary equations: $J_T(x_T, \Gamma_T) = f_T(x_T)$ for any $x_T$ and $\Gamma_T$. Equations (32) and (33) generate convex problems. Although the cost-to-go functions are now two-dimensional, the approach remains tractable because the cost-to-go function at time $t$ for a budget $\Gamma_t$ only depends on the cost-to-go function at time $t+1$ for the budgets $\Gamma_t$ and $\Gamma_t - 1$ (and never for budget values greater than $\Gamma_t$.) Hence, the recursive equations can be solved by a *greedy algorithm* that computes the cost-to-go functions by increasing the second variable from 0 to $\Gamma$ and, for each $\gamma \in \{0, \ldots, \Gamma\}$, decreasing the time period from $T-1$ to 0.

Thiele [47] implements this method in revenue management and derives insights into the impact of uncertainty on the optimal policy. Following the same line of thought, Bienstock and Ozbay [21] provide compelling evidence of the tractability of the approach in the context of inventory management.

## 3.3 Affine and Finite Adaptability

### 3.3.1 Affine Adaptability

Ben-Tal et. al. [6] first extended the robust optimization framework to dynamic settings, where the decision-maker adjusts his strategy to information revealed over time using *policies* rather than re-optimization. Their initial focus was on two-stage decision-making, which in the stochastic programming literature (e.g., Birge and Louveaux [22]) is referred to as optimization with recourse. Ben-Tal et. al. [6] have coined the term "adjustable optimization" for this class of problems when considered in the robust optimization framework. Two-stage problems are characterized by the following sequence of events:

1. the decision-maker selects the "here-and-now", or first-stage, variables, *before* having any knowledge of the actual value taken by the uncertainty,

2. he observes the realizations of the random variables,

3. he chooses the "wait-and-see", or second-stage, variables, *after* learning of the outcome of the random event.

In stochastic programming, the sources of randomness obey a discrete, known distribution and the decision-maker minimizes the sum of the first-stage and the expected second-stage costs. This is for instance justified when the manager can repeat the same experiment a large number of times, has learnt the distribution of the uncertainty in the past through historical data and this distribution does not change. However, such assumptions are rarely satisfied in practice and the decision-maker must then take action with a limited amount of information at his disposal. In that case, an approach based on robust optimization is in order.

The **adjustable robust counterpart** defined by Ben-Tal et. al. [6] ensures feasibility of the constraints for any realizations of the uncertainty, through the appropriate selection of the second-stage decision variables $\mathbf{y}(\omega)$, while minimizing (without loss of generality) a deterministic cost:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}(\omega)} \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b}, \\ & \mathbf{T}(\omega)\,\mathbf{x} + \mathbf{W}(\omega)\mathbf{y}(\omega) \geq \mathbf{h}(\omega), \quad \forall \omega \in \Omega, \end{aligned} \tag{34}$$

where $\{[\mathbf{T}(\omega), \mathbf{W}(\omega), \mathbf{h}(\omega)], \omega \in \Omega\}$ is a convex uncertainty set describing the possible values taken by the uncertain parameters. In contrast, the **robust counterpart** does not allow for the decision variables to depend on the realization of the uncertainty:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b}, \\ & \mathbf{T}(\omega)\,\mathbf{x} + \mathbf{W}(\omega)\mathbf{y} \geq \mathbf{h}(\omega), \quad \forall \omega \in \Omega. \end{aligned} \tag{35}$$

Ben-Tal et. al. [6] show that: (i) Problems (34) and (35) are equivalent in the case of *constraint-wise uncertainty*, i.e., randomness affects each constraint independently, and (ii) in general, Problem (34) is more flexible than Problem (35), but this flexibility comes at the expense of tractability (in mathematical terms, Problem (34) is *NP-hard.*) To address this issue, the authors propose to restrict the second-stage recourse to be an *affine* function of the realized data, i.e., $\mathbf{y}(\omega) = \mathbf{p} + \mathbf{Q}\,\omega$ for some

$\mathbf{p}, \mathbf{Q}$ to be determined. The **affinely adjustable robust counterpart** is defined as:

$$\begin{aligned}\min_{\mathbf{x},\mathbf{p},\mathbf{Q}} \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b}, \\ & \mathbf{T}(\omega)\,\mathbf{x} + \mathbf{W}(\omega)(\mathbf{p} + \mathbf{Q}\,\omega) \geq \mathbf{h}(\omega), \quad \forall \omega \in \Omega.\end{aligned} \tag{36}$$

In many practical applications, and most of the stochastic programming literature, the recourse matrix $\mathbf{W}(\omega)$ is assumed constant, independent of the uncertainty; this case is known as *fixed recourse*. Using strong duality arguments, Ben-Tal et. al. [6] show that Problem (36) can be solved efficiently for special structures of the set $\Omega$, in particular, for polyhedra and ellipsoids. In a related work, Ben-Tal et. al. [5] implement these techniques for retailer-supplier contracts over a finite horizon and perform a large simulation study, with promising numerical results. Two-stage robust optimization has also received attention in application areas such as network design and operation under demand uncertainty (Atamturk and Zhang [3]).

Affine adaptability has the advantage of providing the decision-maker with robust linear *policies*, which are intuitive and relatively easy to implement for well-chosen models of uncertainty. From a theoretical viewpoint, linear decision rules are known to be optimal in *linear-quadratic control*, i.e., control of a system with linear dynamics and quadratic costs (Bertsekas [11]). The main drawback, however, is that there is little justification for the linear decision rule outside this setting. In particular, multi-stage problems in operations research often yield formulations with *linear* costs and linear dynamics, and since quadratic costs lead to linear (or affine) control, it is not unreasonable when costs are linear to expect good performance from *piecewise constant* decision rules. This claim is motivated for instance by results on the optimal control of fluid models (Ricard [37].)

### 3.3.2 Finite Adaptability

The concept of finite adaptability, first proposed by Bertsimas and Caramanis [14], is based on the selection of a finite number of (constant) contingency plans to incorporate the information revealed over time. This can be motivated as follows. While robust optimization is well-suited for problems where uncertainty is aggregated, i.e., constraint-wise, immunizing a problem against uncertainty that cannot be decoupled across constraints yields *overly conservative solutions,* in the sense that the robust approach protects the system against parameters that fall outside the uncertainty set (Soyster [44]). Hence, the decision-maker would benefit from gathering some limited information on the actual value taken by the randomness before implementing a strategy. We focus in this tutorial on two-stage models; the framework also has obvious potential in multi-stage problems.

The recourse under finite adaptability is *piecewise constant* in the number $K$ of contingency plans; therefore, the task of the decision-maker is to partition the uncertainty set into $K$ pieces and determine the best response in each. Appealing features of this approach are that (i) it provides a *hierarchy* of adaptability, and (ii) it is able to incorporate integer second-stage variables and non-convex uncertainty sets, while other proposals of adaptability cannot. We present some of Bertsimas and Caramanis's [14] results below, and in particular, geometric insights into the performance of the $K$-adaptable approach.

*Right-hand side uncertainty*

A robust linear programming problem with right-hand side uncertainty can be formulated as:

$$\begin{aligned} \min \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b}, \quad \forall \mathbf{b} \in \mathcal{B}, \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{37}$$

where $\mathcal{B}$ is the polyhedral uncertainty set for the right-hand-side vector $\mathbf{b}$ and $\mathcal{X}$ is a polyhedron, not subject to uncertainty. To ensure that the constraints $\mathbf{A}\mathbf{x} \geq \mathbf{b}$ hold for all $\mathbf{b} \in \mathcal{B}$, the decision-maker must immunize each constraint $i$ against uncertainty:

$$\mathbf{a_i}'\mathbf{x} \geq b_i, \ \forall \mathbf{b} \in \mathcal{B}, \tag{38}$$

which yields:

$$\mathbf{A}\mathbf{x} \geq \widetilde{\mathbf{b}}_{\mathbf{0}}, \tag{39}$$

where $(\widetilde{b}_0)_i = \max\{b_i \, \text{s.t.} \mathbf{b} \in \mathcal{B}\}$ for all $i$. Therefore, solving the robust problem is equivalent to solving the deterministic problem with the right-hand side being equal to $\widetilde{\mathbf{b}}_{\mathbf{0}}$. Note that $\widetilde{\mathbf{b}}_{\mathbf{0}}$ is the "upper-right" corner of the smallest hypercube $\mathcal{B}_0$ containing $\mathcal{B}$, and might fall far outside the uncertainty set. In that case, non-adjustable robust optimization forces the decision-maker to plan for a very unlikely outcome, which is an obvious drawback to the adoption of the approach by practitioners.

To address the issue of overconservatism, Bertsimas and Caramanis [14] cover the uncertainty set $\mathcal{B}$ with a partition of $K$ (not necessarily disjoint) pieces: $\mathcal{B} = \cup_{k=1}^{K} \mathcal{B}_k$, and select a contingency plan $\mathbf{x_k}$ for each subset $\mathcal{B}_k$. The $K$**-adaptable robust counterpart** is defined as:

$$\begin{aligned} \min \quad & \max_{k=1,\ldots,K} \mathbf{c}'\mathbf{x_k} \\ \text{s.t.} \quad & \mathbf{A}\,\mathbf{x_k} \geq \mathbf{b}, && \forall \mathbf{b} \in \mathcal{B}_k, \ \forall k = 1, \ldots, K, \\ & \mathbf{x_k} \in \mathcal{X}, && \forall k = 1, \ldots, K. \end{aligned} \tag{40}$$

It is straightforward to see that Problem (40) is equivalent to:

$$
\begin{aligned}
\min \quad & \max_{k=1,\ldots,K} \mathbf{c}'\mathbf{x_k} \\
\text{s.t.} \quad & \mathbf{A}\,\mathbf{x_k} \geq \widetilde{\mathbf{b}}_\mathbf{k}, && \forall k = 1,\ldots,K, \\
& \mathbf{x_k} \in \mathcal{X}, && \forall k = 1,\ldots,K,
\end{aligned}
\tag{41}
$$

where $\widetilde{\mathbf{b}}_\mathbf{k}$ is defined as $(\widetilde{b}_k)_i = \max\{b_i \,|\, \mathbf{b} \in \mathcal{B}_k\}$ for each $i$, and represents the upper-right corner of the smallest hypercube containing $\mathcal{B}_k$. Hence, the performance of the finite adaptability approach depends on the choice of the subsets $\mathcal{B}_k$ only through the resulting value of $\widetilde{\mathbf{b}}_\mathbf{k}$, with $k = 1,\ldots,K$. This motivates developing a direct connection between the uncertainty set $\mathcal{B}$ and the vectors $\widetilde{\mathbf{b}}_\mathbf{k}$, without using the subsets $\mathcal{B}_k$.

Let $\mathcal{C}(\mathcal{B})$ be the set of K-uples $(\mathbf{b_1},\ldots,\mathbf{b_K})$ covering the set $\mathcal{B}$, i.e., for any $\mathbf{b} \in \mathcal{B}$, the inequality $\mathbf{b} \leq \mathbf{b_k}$ holds for at least one $k$. The problem of optimally partitioning the uncertainty set into $K$ pieces can be formulated as:

$$
\begin{aligned}
\min \quad & \max_{k=1,\ldots,K} \mathbf{c}'\mathbf{x_k} \\
\text{s.t.} \quad & \mathbf{A}\,\mathbf{x_k} \geq \widetilde{\mathbf{b}}_\mathbf{k}, && \forall k = 1,\ldots,K, \\
& \mathbf{x_k} \in \mathcal{X}, && \forall k = 1,\ldots,K, \\
& (\widetilde{\mathbf{b}}_\mathbf{1},\ldots,\widetilde{\mathbf{b}}_\mathbf{K}) \in \mathcal{C}(\mathcal{B}).
\end{aligned}
\tag{42}
$$

The characterization of $\mathcal{C}(\mathcal{B})$ plays a central role in the approach. Bertsimas and Caramanis [14] investigate in detail the case with two contingency plans, where the decision-maker must select a pair $(\widetilde{\mathbf{b}}_\mathbf{1}, \widetilde{\mathbf{b}}_\mathbf{2})$ that covers the set $\mathcal{B}$. For any $\widetilde{\mathbf{b}}_\mathbf{1}$, the vector $\min(\widetilde{\mathbf{b}}_\mathbf{1}, \widetilde{\mathbf{b}}_\mathbf{0})$ is also feasible and yields a smaller or equal cost in Problem (42). A similar argument holds for $\widetilde{\mathbf{b}}_\mathbf{2}$. Hence, the optimal $(\widetilde{\mathbf{b}}_\mathbf{1}, \widetilde{\mathbf{b}}_\mathbf{2})$ pair in Equation (42) satisfies: $\widetilde{\mathbf{b}}_\mathbf{1} \leq \widetilde{\mathbf{b}}_\mathbf{0}$ and $\widetilde{\mathbf{b}}_\mathbf{2} \leq \widetilde{\mathbf{b}}_\mathbf{0}$. On the other hand, for $(\widetilde{\mathbf{b}}_\mathbf{1}, \widetilde{\mathbf{b}}_\mathbf{2})$ to cover $\mathcal{B}$, we must have either $b_i \leq \widetilde{b}_{1i}$ or $b_i \leq \widetilde{b}_{2i}$ for each component $i$ of any $\mathbf{b} \in \mathcal{B}$. Hence, for each $i$, either $\widetilde{b}_{1i} = \widetilde{b}_{0i}$ or $\widetilde{b}_{2i} = \widehat{b}_{0i}$.

This creates a partition $S$ between the indices $\{1,\ldots,n\}$, where $S = \{i | \widetilde{b}_{1i} = \widetilde{b}_{0i}\}$. $\widetilde{\mathbf{b}}_\mathbf{1}$ is completely characterized by the set $S$, in the sense that $\widetilde{b}_{1i} = \widetilde{b}_{0i}$ for all $i \in S$ and $\widetilde{b}_{1i}$ for $i \notin S$ can be any number smaller than $\widetilde{b}_{0i}$. The part of $\mathcal{B}$ that is not yet covered is $\mathcal{B} \cap \{\exists j,\, b_j \geq \widetilde{b}_{1j}\}$. This forces $\widetilde{b}_{2i} = \widetilde{b}_{0i}$ for all $i \notin S$ and $\widetilde{b}_{2i} \geq \max\{b_i | \mathbf{b} \in \mathcal{B},\, \exists j \in S^c,\, b_j \geq \widetilde{b}_{1j}\}$, or equivalently, $\widetilde{b}_{2i} \geq \max_j \max\{b_i | \mathbf{b} \in \mathcal{B},\, b_j \geq \widetilde{b}_{1j}\}$, for all $i \in S$. Bertsimas and Caramanis [14] show that:

- When $\mathcal{B}$ has a specific structure, the optimal split and corresponding contingency plans can be computed as the solution of a mixed integer-linear program.

- Computing the optimal partition is NP-hard, but can be performed in a tractable manner when either of the following quantities is small: the dimension of the uncertainty, the dimension of the problem, or the number of constraints affected by the uncertainty.

- When none of the quantities above is small, a well-chosen heuristic algorithm exhibits strong empirical performance in large-scale applications.

**Example 3.3 (Newsvendor problem with reorder)** *A manager must order two types of seasonal items before knowing the actual demand for these products. All demand must be met; therefore, once demand is realized, the missing items (if any) are ordered at a more expensive reorder cost. The decision-maker considers two contingency plans. Let $x_j$, $j = 1, 2$ be the amounts of product $j$ ordered before demand is known, and $y_{ij}$ the amount of product $j$ ordered in contingency plan $i$, $i = 1, 2$. We assume that the first-stage ordering costs are equal to 1, and the second-stage ordering costs are equal to 2. Moreover, the uncertainty set for the demand is given by: $\{(d_1, d_2) \,|\, d_1 \geq 0,\, d_2 \geq 0,\, d_1/2 + d_2 \leq 1\}$.*

*The robust, static counterpart would protect the system against $d_1 = 2$, $d_2 = 1$, which falls outside the feasible set, and would yield an optimal cost of 3. To implement the 2-adaptability approach, the decision-maker must select an optimal covering pair $(\widetilde{\mathbf{d_1}}, \widetilde{\mathbf{d_2}})$ satisfying $\widetilde{\mathbf{d_1}} = (d, 1)$ with $0 \leq d \leq 2$ and $\widetilde{\mathbf{d_2}} = (1, d')$ with $d' \geq 1 - d/2$. At optimality, $d' = 1 - d/2$, since increasing the value of $d'$ above that threshold increases the optimal cost while the demand uncertainty set is already completely covered. Hence, the partition is determined by the scalar $d$. Figure 6 depicts the uncertainty set and a possible partition.*

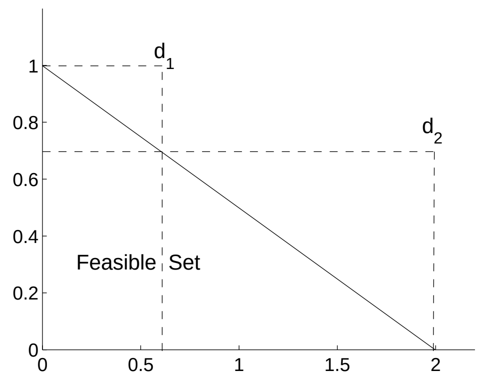


Figure 6: The uncertainty set and a possible partition.

*The 2-adaptable problem can be formulated as:*

$$\begin{aligned} \min \quad & Z \\ s.t. \quad & Z \geq x_1 + x_2 + 2\,(y_{11} + y_{12}), \\ & Z \geq x_1 + x_2 + 2\,(y_{21} + y_{22}), \\ & x_1 + y_{11} \geq d, \\ & x_2 + y_{12} \geq 1, \\ & x_1 + y_{21} \geq 1, \\ & x_2 + y_{22} \geq 1 - d/2, \\ & x_j,\, y_{ij} \geq 0,\, \forall i, j, \\ & 0 \leq d \leq 2. \end{aligned} \tag{43}$$

*The optimal solution is to select* $d = 2/3$, $\mathbf{x} = (2/3,\, 2/3)$ *and* $\mathbf{y_1} = (0, 1/3)$, $\mathbf{y_2} = (1/3, 0)$, *for an optimal cost of* 2. *Hence,* 2*-adaptability achieves a decrease in cost of* 33%.

*Matrix uncertainty*

In this paragraph, we briefly outline Bertsimas and Caramanis's [14] findings in the case of matrix uncertainty and 2-adaptability. For notational convenience, we incorporate constraints without uncertainty ($\mathbf{x} \in \mathcal{X}$ for a given polyhedron $\mathcal{X}$) into the constraints $\mathbf{A}\mathbf{x} \geq \mathbf{b}$. The robust problem can be written as:

$$\begin{aligned} \min \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\,\mathbf{x} \geq \mathbf{b},\ \forall \mathbf{A} \in \mathcal{A}, \end{aligned} \tag{44}$$

where the uncertainty set $\mathcal{A}$ is a polyhedron. Here, we define $\mathcal{A}$ by its extreme points: $\mathcal{A} = conv\{\mathbf{A_1}, \ldots, \mathbf{A_K}\}$, where *conv* denotes the convex hull. Problem (44) becomes:

$$\begin{aligned} \min \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A_k}\mathbf{x} \geq \mathbf{b},\ \forall k = 1, \ldots, K. \end{aligned} \tag{45}$$

Let $\mathcal{A}_0$ be the smallest hypercube containing $\mathcal{A}$. We formulate the 2-adaptability problem as:

$$\begin{aligned} \min \quad & \max\{\mathbf{c}'\mathbf{x_1}, \mathbf{c}'\mathbf{x_2}\} \\ \text{s.t.} \quad & \mathbf{A}\,\mathbf{x_1} \geq \mathbf{b},\ \forall \mathbf{A} \in \mathcal{A}_1, \\ & \mathbf{A}\,\mathbf{x_2} \geq \mathbf{b},\ \forall \mathbf{A} \in \mathcal{A}_2, \end{aligned} \tag{46}$$

where $\mathcal{A} \subset (\mathcal{A}_1 \cup \mathcal{A}_2) \subset \mathcal{A}_0$.

Bertsimas and Caramanis [14] investigate in detail the conditions for which the 2-adaptable approach improves the cost of the robust static solution by at least $\eta > 0$. Let $\mathbf{A_0}$ be the corner point of $\mathcal{A}_0$ such that Problem (44) is equivalent to $\min\ \mathbf{c}'\mathbf{x}\,\text{s.t.}\,\mathbf{A_0}\,\mathbf{x} \geq \mathbf{b}$. Intuitively, the decision-maker needs to remove from the partition $\mathcal{A}_1 \cup \mathcal{A}_2$ an area around $\mathbf{A_0}$ large enough to ensure this cost decrease. The authors build upon this insight to provide a geometric perspective on the gap between the robust and the 2-adaptable frameworks. A key insight is that, if $v^*$ is the optimal objective of the robust problem (44), the problem:

$$\begin{aligned} \min\ & 0 \\ \text{s.t}\ & \mathbf{A^i}\,\mathbf{x} \geq \mathbf{b},\ \forall i = 1, \ldots, K, \\ & \mathbf{c}'\mathbf{x} \leq v^* - \eta \end{aligned} \tag{47}$$

is infeasible. Its dual is feasible (for instance, $\mathbf{0}$ belongs to the feasible set) and hence unbounded by strong duality. The set $\mathcal{D}$ of directions of dual unboundedness is obtained by scaling the extreme rays:

$$\mathcal{D} = \left\{ (\mathbf{p_1}, \ldots, \mathbf{p_K}) \,|\, \mathbf{b}' \left( \sum_{i=1}^{K} \mathbf{p_i} \right) \geq v^* - \eta,\ \sum_{i=1}^{K} (\mathbf{A^i})'\mathbf{p_i} = \mathbf{c},\ \mathbf{p_1}, \ldots, \mathbf{p_K} \geq \mathbf{0}. \right\}. \tag{48}$$

The $(\mathbf{p_1}, \ldots, \mathbf{p_K})$ in the set $\mathcal{D}$ are used to construct a family $\mathcal{A}_\eta$ of matrices $\widetilde{\mathbf{A}}$ such that the optimal cost of the nominal problem (solved for any matrix in this family) is at least equal to $v^* - \eta$. (This is simply done by defining $\widetilde{\mathbf{A}}$ such that $\sum_{i=1}^{K} \mathbf{p_i}$ is feasible for the dual of the nominal problem, i.e., $\widetilde{\mathbf{A}}' \sum_{i=1}^{K} \mathbf{p_i} = \sum_{i=1}^{K} (\mathbf{A^i})'\mathbf{p_i}$.) The family $\mathcal{A}_\eta$ plays a crucial role in understanding the performance of the 2-adaptable approach. Specifically, 2-adaptability decreases the cost by strictly more than $\eta$ if and only if $\mathcal{A}_\eta$ has no element in the partition $\mathcal{A}_1 \cup \mathcal{A}_2$. The reader is referred to [14] for additional properties.

As pointed out in [14], finite adaptability is *complementary* to the concept of affinely adjustable optimization proposed by Ben-Tal et. al. [6], in the sense that neither technique performs consistently better than the other. Understanding the problem structure required for good performance of these techniques is an important future research direction. Bertsimas et. al. [15] apply the adaptable framework to air traffic control subject to weather uncertainty, where they demonstrate the method's ability to incorporate randomness in very large-scale integer formulations.

# 4 Connection with Risk Preferences

## 4.1 Robust optimization and coherent risk measures

So far, we have assumed that the polyhedral set describing the uncertainty was *given*, and developed robust optimization models based on that input. In practice however, the true information available to the decision-maker is historical data, which must be incorporated into an uncertainty set before the robust optimization approach can be implemented. We now present an explicit methodology to construct this set, based on past observations of the random variables and the decision-maker's attitude towards risk. The approach is due to Bertsimas and Brown [13]. An application of data-driven optimization to inventory management is presented in Bertsimas and Thiele [19].

We consider the following problem:

$$\begin{aligned} \min \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}'\mathbf{x} \leq b, \\ & \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{49}$$

The decision-maker has $N$ historical observations $\mathbf{a_1}, \ldots, \mathbf{a_N}$ of the random vector $\tilde{\mathbf{a}}$ at his disposal. Therefore, for any given $\mathbf{x}$, $\tilde{\mathbf{a}}'\mathbf{x}$ is a random variable whose sample distribution is given by: $P[\tilde{\mathbf{a}}'\mathbf{x} = \mathbf{a_i}'\mathbf{x}] = 1/N$, for $i = 1, \ldots, N$. (We assume that the $\mathbf{a_i}'\mathbf{x}$ are distinct, the extension to the general case is straightforward.) The decision-maker associates a numerical value $\mu(\tilde{\mathbf{a}}'\mathbf{x})$ to the random variable $\tilde{\mathbf{a}}'\mathbf{x}$; the function $\mu$ captures his attitude towards risk and is called a *risk measure*. We then define the **risk-averse problem** as:

$$\begin{aligned} \min \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mu(\tilde{\mathbf{a}}'\mathbf{x}) \leq b, \\ & \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{50}$$

While any function from the space of almost surely bounded random variables $\mathcal{S}$ to the space of real numbers $\mathcal{R}$ can be selected as a risk measure, some are more sensible choices than others. In particular, Artzner et. al. [1] argue that a measure of risk should satisfy four axioms, which define the class of *coherent risk measures*:

1. Translation invariance: $\mu(X + a) = \mu(X) - a, \ \forall X \in \mathcal{S}, \, a \in \mathcal{R}$.

2. Monotonicity: if $X \leq Y$ w.p. 1, $\mu(X) \leq \mu(Y), \ \forall X, Y \in \mathcal{S}$.

3. Subadditivity: $\mu(X + Y) \leq \mu(X) + \mu(Y), \ \forall X, Y \in \mathcal{S}$.

4. Positive homogeneity: $\mu(\lambda X) = \lambda \, \mu(X), \ \forall X \in \mathcal{S}, \, \lambda \geq 0$.

An example of a coherent risk measure is the tail conditional expectation, i.e., the expected value of the losses given that they exceed some quantile. Other risk measures such as standard deviation and the probability that losses will exceed a threshold, also known as Value-at-Risk, are not coherent for general probability distributions.

An important property of coherent risk measures is that they can be represented as the *worst-case expected value over a family of distributions.* Specifically, $\mu$ is coherent if and only if there exists a family of probability measures $\mathcal{Q}$ such that:

$$\mu(X) = \sup_{q \in \mathcal{Q}} E_q[X], \ \forall X \in \mathcal{S}. \tag{51}$$

In particular, if $\mu$ is a coherent risk measure and $\tilde{\mathbf{a}}$ is distributed according to its sample distribution ($P(\mathbf{a} = \mathbf{a_i}) = 1/N$ for all $i$), Bertsimas and Brown [13] note that:

$$\mu(\tilde{\mathbf{a}}'\mathbf{x}) = \sup_{q \in \mathcal{Q}} E_Q[\tilde{\mathbf{a}}'\mathbf{x}] = \sup_{q \in \mathcal{Q}} \sum_{i=1}^{N} q_i \, \mathbf{a_i}'\mathbf{x} = \sup_{a \in \mathcal{A}} \mathbf{a}'\mathbf{x}, \tag{52}$$

with the uncertainty set $\mathcal{A}$ defined by:

$$\mathcal{A} = conv \left\{ \sum_{i=1}^{N} q_i \, \mathbf{a_i} | \mathbf{q} \in \mathcal{Q} \right\}, \tag{53}$$

and the risk-averse problem (50) is then *equivalent* to the robust optimization problem:

$$\begin{aligned} \min \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}'\mathbf{x} \leq b, \ \forall \mathbf{a} \in \mathcal{A}, \\ & \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{54}$$

The convex (not necessarily polyhedral) uncertainty set $\mathcal{A}$ is included into the convex hull of the data points $\mathbf{a_1}, \ldots, \mathbf{a_N}$. Equation (53) provides an *explicit characterization* of the uncertainty set that the decision-maker should use if his attitude towards risk is based on a coherent risk measure. It also raises two questions: (i) can we obtain the generating family $\mathcal{Q}$ easily, at least for some well-chosen coherent risk measures? (ii) can we identify risk measures that lead to polyhedral uncertainty sets, since those sets have been central to the robust optimization approach presented so far? In Section 4.2, we address both issues simultaneously by introducing the concept of *comonotone risk measures.*

## 4.2 Comonotone risk measures

To investigate the connection between the decision-maker's attitude towards risk and the choice of polyhedral uncertainty sets, Bertsimas and Brown [13] consider a second representation of coherent risk measures based on *Choquet integrals.*

The Choquet integral $\mu_g$ of a random variable $X \in \mathcal{S}$ with respect to the distortion function $g$ (which can be any non-decreasing function on $[0, 1]$ such that $g(0) = 0$ and $g(1) = 1$) is defined by:

$$\mu_g(X) = \int_0^\infty g(P[X \geq x]) \, dx + \int_{-\infty}^0 [g(P[X \geq x]) - 1] \, dx. \tag{55}$$

$\mu_g$ is coherent if and only if $g$ is concave (Reesor and McLeish [36]). While not every coherent risk measure can be re-cast as the expected value of a random variable under a distortion function, Choquet integrals provide a broad modeling framework, which includes conditional tail expectation and value-at-risk. Schmeidler [39] shows that a risk measure can be represented as a Choquet integral with a concave distortion function (and hence be coherent) if and only if the risk measure satisfies a property called *comonotonicity.*

A random variable is said to be comonotonic if its support $S$ has a complete order structure (for any $\mathbf{x}, \mathbf{y} \in S$, either $\mathbf{x} \leq \mathbf{y}$ or $\mathbf{y} \leq \mathbf{x}$), and a risk measure is said to be comonotone if for any comonotonic random variables $X$ and $Y$, we have:

$$\mu(X+Y) = \mu(X) + \mu(Y). \tag{56}$$

**Example 4.1 (Comonotonic random variable, Bertsimas and Brown [13])**
*Consider the joint payoff of a stock and a call option on that stock. With $S$ the stock value and $K$ the strike price of the call option, the joint payoff $(S, \max(0, S-K))$ is obviously comonotonic. For instance, with $K = 2$ and $S$ taking any value between 1 and 5, the joint payoff takes values $\mathbf{x_1} = (1, 0)$, $\mathbf{x_2} = (2, 0)$, $\mathbf{x_3} = (3, 1)$, $\mathbf{x_4} = (4, 2)$ and $\mathbf{x_5} = (5, 3)$. Hence, $\mathbf{x_{i+1}} \geq \mathbf{x_i}$ for each $i$.*

Bertsimas and Brown [13] show that, for any comonotone risk measure with distortion function $g$, noted $\mu_g$, and any random variable $Y$ with support $\{y_1, \ldots, y_N\}$ such that $P[Y = y_i] = 1/N$, $\mu_g$ can be computed using the formula:

$$\mu_g(Y) = \sum_{i=1}^{N} q_i \, y_{(i)}, \tag{57}$$

where $y_{(i)}$ is the $i$-th smallest $y_j$, $j = 1, \ldots, N$ (hence, $y_{(1)} \leq \ldots \leq y_{(N)}$), and $q_i$ is defined by:

$$q_i = g\left(\frac{N+1-i}{N}\right) - g\left(\frac{N-i}{N}\right). \tag{58}$$

Because $g$ is non-decreasing and concave, it is easy to see that the $q_i$ are non-decreasing. Bertsimas and Brown [13] use this insight to represent $\sum_{i=1}^{N} q_i \, y_{(i)}$ as the optimal solution of a linear programming problem:

$$
\begin{aligned}
\max \quad & \sum_{i=1}^{N}\sum_{j=1}^{N} q_i\, y_j\, w_{ij} \\
\text{s.t.} \quad & \sum_{i=1}^{N} w_{ij} = 1, \qquad \forall j, \\
& \sum_{j=1}^{N} w_{ij} = 1, \qquad \forall i, \\
& w_{ij} \geq 0, \qquad \forall i, j.
\end{aligned}
\tag{59}
$$

At optimality the largest $y_i$ is assigned to $q_N$, the second largest to $q_{N-1}$, and so on. Let $W(N)$ be the feasible set of Problem (59). Equation (57) becomes:

$$
\mu_g(Y) = \max_{\mathbf{w}\in W(N)} \sum_{i=1}^{N}\sum_{j=1}^{N} q_i\, y_j\, w_{ij}. \tag{60}
$$

This yields a generating family $\mathcal{Q}$ for $\mu_g$:

$$
\mathcal{Q} = \left\{\mathbf{w}'\mathbf{q},\, \mathbf{w} \in W(N)\right\}, \tag{61}
$$

or equivalently, using the optimal value of $\mathbf{w}$:

$$
\mathcal{Q} = \{\mathbf{p}, \exists \sigma \in S_N,\, p_i = q_{\sigma(i)},\, \forall i\}, \tag{62}
$$

where $S_N$ is the group of permutations over $\{1, \ldots, N\}$. Bertsimas and Brown [13] make the following observations:

- While coherent risk measures are in general defined by a *family* $\mathcal{Q}$ of probability distributions, comonotone risk measures require the knowledge of a *single generating vector* $\mathbf{q}$. The family $\mathcal{Q}$ is then derived according to Equation (62).
- Comonotone risk measures lead to polyhedral uncertainty sets of a *specific structure*: the convex hull of all $N!$ convex combinations of $\{\mathbf{a_1}, \ldots, \mathbf{a_N}\}$ induced by all permutations of the vector $\mathbf{q}$.

It follows from injecting the generating family $\mathcal{Q}$ given by Equation (62) into the definition of the uncertainty set $\mathcal{A}$ in Equation (53) that the risk-averse problem (50) is equivalent to the robust optimization problem solved for the polyhedral uncertainty set:

$$
\mathcal{A}_q = conv\left\{\sum_{i=1}^{N} q_{\sigma(i)}\, a_i,\, \sigma \in S_N\right\}. \tag{63}
$$

Note that $\mathbf{q} = (1/N)\,\mathbf{e}$ with $\mathbf{e}$ the vector of all one's yields the sample average $(1/N)\sum_{i=1}^{N} \mathbf{a_i}$ and $\mathbf{q} = (1, 0, \ldots, 0)$ yields the convex hull of the data. Figure 7 shows possible uncertainty sets with $N = 5$ observations.

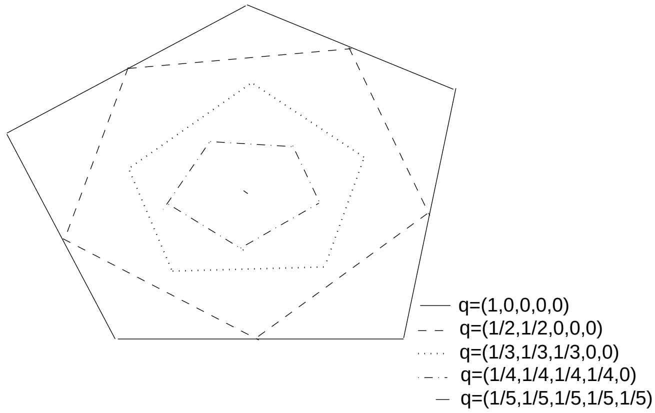


Figure 7: Uncertainty sets derived from comonotone risk measures.

### 4.3 Additional results

Bertsimas and Brown [13] provide a number of additional results connecting coherent risk measures and convex uncertainty sets. We enumerate a few here:

1. Tail conditional expectations $CTE_{i/N}$, $i = 1, \ldots, N$, can be interpreted as *basis functions* for the entire space of comonotone risk measures on random variables with a discrete state space of size $N$.

2. The class of *symmetric polyhedral uncertainty sets* is generated by a specific set of coherent risk measures. These uncertainty sets are useful because they naturally induce a norm.

3. Optimization over the following coherent risk measure based on higher-order tail moments:

$$\mu_{p,\alpha}(X) = E[X] + \alpha\,(E\,[(\max\{0, X - E[X]\})^p])])^{1/p} \qquad (64)$$

   is equivalent to a robust optimization problem with a norm-bounded uncertainty set.

4. Any robust optimization problem with a convex uncertainty set (contained within the convex hull of the data) can be reformulated as a risk-averse problem with a coherent risk measure.

## 5 Conclusions

Robust optimization has emerged over the last decade as a tractable, insightful approach to decision-making under uncertainty. It is well-suited for both static and dynamic problems with imprecise information, has a strong connection with the decision-maker's attitude towards risk and can be applied in numerous areas, including inventory management, air traffic control, revenue management, network design and portfolio optimization. While this tutorial has primarily focused on linear programming and polyhedral uncertainty sets, the modeling power of robust optimization extends to more general settings, for instance second-order cone programming and ellipsoidal uncertainty sets. It has also been successfully implemented in stochastic and dynamic programming with ambiguous probabilities. Current topics of interest include: (i) tractable methods to incorporate information revealed over time in multi-stage problems, and (ii) data-driven optimization, which injects historical data directly into the mathematical programming model, for instance through explicit guidelines to construct the uncertainty set. Hence, the robust and data-driven framework provides a compelling alternative to traditional decision-making techniques under uncertainty.

## References


[1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[2] A. Atamturk. Strong formulations of robust mixed 0-1 programming. *Mathematical Programming*, 2005.

[3] A. Atamturk and M. Zhang. Two-stage robust network flow and design under demand uncertainty. Technical report, University of California at Berkeley, 2004.

[4] A. Ben-Tal, S. Boyd, and A. Nemirovski. Extending the scope of robust optimization: comprehensive robust counterparts of uncertain problems. Technical report, Georgia Institute of Technology, 2005.

[5] A. Ben-Tal, B. Golani, A. Nemirovski, and J.-P. Vial. Supplier-retailer flexible commitments contracts: a robust optimization approach. *Manufacturing and Service Operations Management*, 7(3):248–273, 2005.

[6] A. Ben-Tal, A. Goryashko, E. Guslitser, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99:351–376, 2004.

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

[8] A. Ben-Tal and A. Nemirovski. Robust solutions to uncertain programs. *Operations Research Letters*, 25:1–13, 1999.

[9] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88:411–424, 2000.

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

[11] D. Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena Scientific, 2nd edition, 2001.

[12] D. Bertsekas and J. Tsitsiklis. *Neuro-dynamic programming.* Athena Scientific, 1996.

[13] D. Bertsimas and D. Brown. Robust linear optimization and coherent risk measures. Technical report, Massachusetts Institute of Technology, 2005.

[14] D. Bertsimas and C. Caramanis. Finite adaptability in linear optimization. Technical report, Massachusetts Institute of Technology, 2005.

[15] D. Bertsimas, C. Caramanis, and W. Moser. Multistage finite adaptability: Application to air traffic control. Working Paper, Massachusetts Institute of Technology, 2006.

[16] D. Bertsimas, D. Pachamanova, and M. Sim. Robust linear optimization under general norms. *Operations Research Letters*, 32(6):510–516, 2004.

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

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

[19] D. Bertsimas and A. Thiele. A data-driven approach to newsvendor problems. Technical report, Massachusetts Institute of Technology, 2004.

[20] D. Bertsimas and A. Thiele. A robust optimization approach to inventory theory. *Operations Research*, 54(1):150–168, 2006.

[21] D. Bienstock and N. Ozbay. Computing optimal basestocks. Technical report, Columbia University, 2005.

[22] J. Birge and F. Louveaux. *Introduction to stochastic programming.* Springer Verlag, 1997.

[23] A. Charnes and W. Cooper. Chance-constrained programming. *Management Science*, 6(1):73–79, 1959.

[24] A. Clark and H. Scarf. Optimal policies for a multi-echelon inventory problem. *Management Science*, 6(4):475–490, 1960.

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

[26] J. Dupacova. The minimax approach to stochastic programming and an illustrative application. *Stochastics*, 20:73–88, 1987.

[27] L. El-Ghaoui and H. Lebret. Robust solutions to least-square problems to uncertain data matrices. *SIAM Journal on Matrix Analysis and Applications*, 18:1035–1064, 1997.

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

[29] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.

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

[31] P. Kall and J. Mayer. *Stochastic linear programming: Models, theory and computation.* Springer Verlag, 2005.

[32] S. Nahmias. *Production and operations analysis.* McGraw-Hill, 5th edition, 2005.

[33] A. Nilim and L. El-Ghaoui. Robust control of markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005.

[34] F. Ordonez and J. Zhao. Robust capacity expansion of network flows. Technical report, University of Southern California, 2005.

[35] E. Porteus. *Foundations of stochastic inventory theory.* Stanford University Press, 2002.

[36] M. Reesor and D. McLeish. Risk, entropy and the transformation of distributions. Technical report, Bank of Canada, 2002.

[37] M. Ricard. *Optimization of queueing networks, an optimal control approach.* PhD thesis, Massachusetts Institute of Technology, 1995.

[38] H. Scarf. *Studies in the mathematical theory of inventory and production*, chapter A min-max solution of an inventory problem, pages 201–209. Stanford University Press, 1958.

[39] D. Schmeidler. Integral representation without additivity. In *Proc. of the American Mathematical Society*, 1986.

[40] A. Shapiro. Worst-case distribution analysis of stochastic programs. *Mathematical Programming*, 2005.

[41] Y. Sheffi. *The resilient enterprise: overcoming vulnerability for competitive advantage.* MIT Press, 2005.

[42] M. Sim. *Robust optimization.* PhD thesis, Massachusetts Institute of Technology, 2004.

[43] D. Simchi-Levi, P. Kaminsky, and E. Simchi-Levi. *Managing the supply chain: the definitive guide for the business professional.* McGraw-Hill, 2004.

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

[45] A. Thiele. *A robust optimization approach to supply chain and revenue management.* PhD thesis, Massachusetts Institute of Technology, 2004.

[46] A. Thiele. Robust dynamic optimization: a distribution-free approach. Technical report, Lehigh University, 2005.

[47] A. Thiele. Robust revenue management with dynamic budgets of uncertainty. Technical report, Lehigh University, 2005.

[48] J. Zackova. On minimax solutions of stochastic linear programming problems. *Cas. Pest. Mat.*, 91:423–430, 1966.