The Decision Rule Approach to Optimization under Uncertainty: Methodology and Applications

Angelos Georghiou[1], Daniel Kuhn[2], and Wolfram Wiesemann[3]

[1] *Desautels Faculty of Management, McGill University, Canada*

[2] *Risk Analytics and Optimization Chair, EPFL, Lausanne, Switzerland*

[3] *Imperial College Business School, Imperial College London, United Kingdom*

Abstract

Dynamic decision-making under uncertainty has a long and distinguished history in operations research. Due to the curse of dimensionality, solution schemes that naïvely partition or discretize the support of the random problem parameters are limited to small and medium-sized problems, or they require restrictive modeling assumptions (*e.g.*, absence of recourse actions). In the last few decades, several solution techniques have been proposed that aim to alleviate the curse of dimensionality. Amongst these is the *decision rule approach*, which faithfully models the random process and instead approximates the feasible region of the decision problem. In this paper, we survey the major theoretical findings relating to this approach, and we investigate its potential in two applications areas.

Keywords. Robust Optimization, Stochastic Programming, Decision Rules, Optimization under Uncertainty.

1 Introduction

Operations managers frequently take decisions whose consequences extend well into the future. Inevitably, the outcomes of such choices are affected by significant uncertainty: changes in customer taste, technological advances and unforeseen stakeholder actions all have a bearing on the suitability of these decisions. It is well documented in theory and practice that disregarding this uncertainty often results in severely suboptimal decisions, which can in turn lead to the underperformance or complete breakdown of production processes. Yet, researchers and practitioners frequently neglect uncertainty and instead focus on the expected or most likely market developments. At first glance, this simplified view appears to be justified by the curse of dimensionality, which plagues dynamic optimization problems under uncertainty. In this paper, we argue that modern approximation schemes in stochastic and robust optimization offer an attractive tradeoff between optimality and tractability, and they are mature enough to be used in practical applications. In particular, we will review the *decision rule approach*, which has a long history in stochastic programming [24, 25] and which has recently been rediscovered by the (distributionally) robust optimization community [8].

Dynamic decision problems under uncertainty have been studied, amongst others, by the stochastic programming and the robust optimization communities. Stochastic programs model the uncertain parameters of a decision problem as a random vector that follows a known distribution. This distribution is typically approximated to gain tractability. While such discretization schemes work well for problems with two time stages, especially when combined with a Bender's decomposition [22], naïve extensions of such schemes to multiple time stages suffer from the curse of dimensionality. To alleviate this problem, the stochastic programming literature reports sophisticated discretization schemes that allow to faithfully approximate the dynamic decision problem in a tractable way [49, 65, 66, 67, 72]. Alternatively, if the stochastic process possesses the Markov property (*i.e.*, the random vectors of different time stages are pairwise independent), one can use stochastic dual dynamic programming [63, 71] to ease the computational burden.

Traditionally, robust optimization replaces probability distributions with uncertainty sets as fundamental building blocks. The goal is to determine a policy that performs best in view of the most adverse parameter realization from within the uncertainty set. Two-stage robust optimization problems are often solved with adaptations of Bender's decomposition [5, 84, 73]. Alternatively, two-stage robust optimization problems can be conservatively approximated by their K-adaptability formulations, which select K candidate second-stage decisions here-and-now (*i.e.*, before the realization of the uncertain parameter vector is observed) and implement the best of these decisions after the realization is observed [11, 17, 48, 74]. Two-stage robust optimization problems can also be formulated as copositive programs, which can be conservatively approximated via semidefinite programming [47, 82]. Multi-stage robust optimization problems with $T > 2$ time stages, on the other hand, are typically approximated conservatively through decision rules, which restrict the admissible recourse actions to affine [44, 53], segregated affine [27, 28, 41], piecewise affine [38, 39] and algebraic as well as trigonometric polynomial [19] functions of the observed parameter values. Decision rules have recently been extended to incorporate both continuous and discrete second-stage decisions, either by partitioning the uncertainty set into hyperrectangles [14, 42, 78] or by resorting to a semi-infinite solution scheme [13]. By themselves, decision rule approximations only provide a conservative bound on the optimal value of the multi-stage problem. To estimate the incurred suboptimality, decision rules are often combined with progressive bounds that emerge from replacing the uncertainty set with a finite subset of the parameter realizations. Scenario subsets that lead to good progressive bounds can be obtained from the Lagrange multipliers associated with the optimal solution of the decision rule problem [12, 46]. The suboptimality of decision rules in the multi-stage setting has been investigated in [53]. As an alternative to decision rules, a robust variant of stochastic dual dynamic programming has recently been proposed in [37].

Stochastic programming and robust optimization have been successfully employed in diverse application domains, ranging from network flow problems [4] and vehicle routing [43, 75] to railway shunting and timetabling [55], energy systems [57, 58, 69, 81], the strategic [2] and operative [3, 21, 54] aspects of operations management as well as healthcare [6, 51]. For a detailed review of these applications, we refer the reader to [10, 35, 83].

The purpose of this paper is to provide an introductory survey of the decision rule approach, complementing the recent review article [31]. To date, the theoretical developments of the decision rule approach are scattered over several papers, and due to their technical terminology they have attracted little attention outside the mathematical optimization community. We aim to present the key findings of this field in a comprehensible and unified framework. To demonstrate the effectiveness of the decision rule approach, we apply the method to two stylized case studies in production planning and supply chain design. The case studies also demonstrate the value of a faithful modeling of uncertainty in multi-stage decision-making.

The remainder of the paper is structured as follows. In Section 2, we formulate the decision problems that we are interested in. We introduce linear decision rules in Section 3, and we generalize the approach to non-linear decision rules in Section 4. Section 5 describes an extension to integer here-and-now decisions. We close with two operations management case studies in Section 6.

Notation For a square matrix $A \in \mathbb{R}^{n\times n}$ we denote by $\mathrm{Tr}\,(A)$ the trace of A, that is, the sum of its diagonal entries. For $A, B \in \mathbb{R}^{m\times n}$ the inequalities $A \leq B$ and $A \geq B$ are understood to hold component-wise, and $A^\top$ denotes the transpose of A. For any real number c we define $c^+ = \max\{c, 0\}$. We define $\mathrm{conv}\, X$ as the convex hull of a set X.

2 Decision-Making under Uncertainty

2.1 Problem Formulation

We study dynamic decision problems under uncertainty of the following general structure. A decision maker first observes an uncertain parameter $\xi_1 \in \mathbb{R}^{k_1}$ and then takes a decision $x_1(\xi_1) \in \mathbb{R}^{n_1}$. Subsequently, a second uncertain parameter $\xi_2 \in \mathbb{R}^{k_2}$ is revealed, in response to which the decision maker takes a second decision $x_2(\xi_1, \xi_2) \in \mathbb{R}^{n_2}$. This sequence of alternating observations and decisions extends over T stages, where at any stage $t = 1, \ldots, T$ the decision maker observes an uncertain parameter ξ_t and selects a decision $x_t(\xi_1, \ldots, \xi_t)$. We emphasize that a decision taken at stage t depends on the whole history of past observations $\xi_1, \ldots, \xi_t$, but it may *not* depend on the future observations $\xi_{t+1}, \ldots, \xi_T$. This feature reflects the *non-anticipative* nature of the dynamic decision problem and ensures its causality. To simplify notation, we define the history of observations up to time t as $\xi^t = (\xi_1, \ldots, \xi_t) \in \mathbb{R}^{k^t}$, where $k^t = \sum_{s=1}^{t} k_s$. Moreover, we let $\xi = (\xi_1, \ldots, \xi_T) \in \mathbb{R}^k$ denote the vector concatenation of *all* uncertain parameters, where $k = k^T$.

We assume that the decision taken at stage t incurs a linear cost $c_t(\xi^t)^\top x_t(\xi^t)$, where the vector of cost coefficients depends linearly on the observation history, that is, $c_t(\xi^t) = C_t \xi^t$ for some matrix $C_t \in \mathbb{R}^{n_t \times k^t}$. We also assume that the decisions are required to satisfy a set of linear inequality constraints to be detailed below. The decision maker's objective is to select the functions or *decision rules* $x_1(\cdot), \ldots, x_T(\cdot)$, which map observation histories to decisions, such that the expected total cost is minimized while all inequality constraints are satisfied. Formally, this decision problem can be represented as follows.

$$
\begin{aligned}
&\text{minimize} \quad \mathbb{E}_\xi\left(\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)\right) \\
&\text{subject to} \quad \left.\begin{aligned} \mathbb{E}_\xi\left(\sum_{s=1}^{T} A_{ts} x_s(\xi^s) \;\middle|\; \xi^t\right) \geq b_t(\xi^t) \\ x_t(\xi^t) \geq 0 \end{aligned}\right\} \; \forall \xi \in \Xi,\ t = 1, \ldots, T
\end{aligned}
\tag{$\mathcal{P}$}
$$

Here, $\mathbb{E}_\xi(\cdot)$ denotes expectation with respect to the random parameter ξ, while Ξ stands for the range of all possible values that ξ can adopt. Below, we will refer to Ξ as the *uncertainty set* and to any $\xi \in \Xi$ as a *scenario*. We will henceforth assume that Ξ is a bounded polyhedron of the form $\Xi = \{\xi \in \mathbb{R}^k : W\xi \geq h\}$ for some $W \in \mathbb{R}^{l \times k}$ and $h \in \mathbb{R}^l$. Recall that the stage-t decisions $x_t(\xi^t)$ in problem $\mathcal{P}$ are functions of ξ^t. Intuitively, this means that every observation history ξ^t corresponding to some scenario $\xi \in \Xi$ gives rise to n_t ordinary decision variables. Since the polyhedron Ξ typically contains infinitely many scenarios, problem $\mathcal{P}$ accommodates in fact infinitely many decision variables.

The inequality constraints in problem $\mathcal{P}$ are expressed in terms of deterministic constraint matrices $A_{ts} \in \mathbb{R}^{m_t \times n_s}$ and uncertainty-affected right-hand side vectors $b_t(\xi^t) \in \mathbb{R}^{m_t}$. We assume that $b_t(\xi^t) = B_t \xi^t$ for some matrices $B_t \in \mathbb{R}^{m_t \times k^t}$. We remark that the assumed linearity of $c_t(\xi^t)$ and $b_t(\xi^t)$ in ξ^t is non-restrictive because we are free to redefine ξ such that it contains $c_t(\xi^t)$ and $b_t(\xi^t)$ as subvectors. The assumption that Ξ is a polyhedron is restrictive. However, all results of this paper naturally extend to convex uncertainty sets characterized through conic constraints (see, e.g., [39, 45]). For ease of exposition, we focus on polyhedral uncertainty sets in this survey.

Note that the stage-t constraints in $\mathcal{P}$ are conditioned on the stage-t observation history ξ^t, where $\mathbb{E}_\xi(\cdot|\xi^t)$ denotes the *conditional* expectation with respect to ξ given ξ^t. Hence, $\mathbb{E}_\xi(\cdot|\xi^t)$ treats $\xi_1, \ldots, \xi_t$ as deterministic variables and takes the expectation only with respect to the future observations $\xi_{t+1}, \ldots, \xi_T$. This implies that the stage-t constraints are parameterized in ξ^t in a similar fashion like the stage-t decisions. Indeed, every ξ^t corresponding to some scenario $\xi \in \Xi$ gives rise to m_t ordinary linear constraints. Since Ξ has typically infinite cardinality, problem $\mathcal{P}$ thus accommodates infinitely many constraints. Intuitively, problem $\mathcal{P}$ can therefore be viewed as an infinite-dimensional generalization of the standard linear program. We remark that conditional expectation constraints are somewhat non-standard in the literature. However, as we will see in

Section 3.2, conditional expectation constraints naturally appear in the dual of problem $\mathcal{P}$, which will play a central role in assessing the quality of decision rule approximations. For symmetry, it is thus convenient to account for conditional expectation constraints already in $\mathcal{P}$.

2.2 Expressiveness of Problem $\mathcal{P}$

The general decision problem under uncertainty described in Section 2.1 provides considerable modeling flexibility. Indeed, as we will show in this section, problem $\mathcal{P}$ encapsulates conventional deterministic and stochastic linear programs, robust optimization problems and tight convex approximations of chance-constrained programs as special cases.

Deterministic linear programs If the uncertainty set contains only one single scenario, that is, if $\Xi = \{\xi_*\}$, then problem $\mathcal{P}$ reduces to a deterministic linear program. In this case only the decisions and constraints corresponding to $\xi = \xi_*$ are relevant, while all (conditional and unconditional) expectations become redundant and can thus be eliminated. Introducing the finite problem data

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1T} \\ \vdots & \ddots & \vdots \\ A_{T1} & \cdots & A_{TT} \end{pmatrix}, \quad b = \begin{pmatrix} b_1(\xi_*^1) \\ \vdots \\ b_T(\xi_*^T) \end{pmatrix} \quad \text{and} \quad c = \begin{pmatrix} c_1(\xi_*^1) \\ \vdots \\ c_T(\xi_*^T) \end{pmatrix},$$

we can reformulate $\mathcal{P}$ as the standard linear program

$$\begin{aligned} &\text{minimize} && c^\top x \\ &\text{subject to} && Ax \geq b,\ x \geq 0, \end{aligned} \tag{$\mathcal{LP}$}$$

whose finite-dimensional decision vector can be identified with $(x_1(\xi_*^1), \ldots, x_T(\xi_*^T))$.

Remark 2.1 (Deterministic Decisions and Constraints)**.** *Deterministic (here-and-now) decisions and constraints can conveniently be incorporated into the general problem $\mathcal{P}$ by requiring that ξ_1 is equal to 1 for all $\xi \in \Xi$. This can always be enforced by appending the equation $\xi_1 = 1$ to the definition of Ξ. From now on, we will always assume that $k_1 = 1$ and that any $\xi = (\xi_1, \ldots, \xi_T) \in \Xi$ satisfies $\xi_1 = 1$.*

Stochastic programs Problem $\mathcal{P}$ can be specialized to a standard linear multi-stage stochastic program with recourse if we ensure that the stage-t constraints are not affected by the future decisions $x_{t+1}(\xi^{t+1}), \ldots, x_T(\xi^T)$. This is achieved by setting $A_{ts} = 0$ for all $t < s$ and has the effect that the term inside the conditional expectation of the stage-t constraint becomes independent of $\xi_{t+1}, \ldots, \xi_T$. Since $\mathbb{E}_\xi(\cdot|\xi^t)$ treats ξ^t as a constant, the conditional expectation thus becomes redundant and can be omitted. Therefore, problem $\mathcal{P}$ reduces to the following multi-stage stochastic program in standard form [22, 50].

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_\xi \left(\displaystyle\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t) \right) \\ \text{subject to} & \left. \begin{array}{r} \displaystyle\sum_{s=1}^{t} A_{ts} x_s(\xi^s) \geq b_t(\xi^t) \\ x_t(\xi^t) \geq 0 \end{array} \right\} \forall \xi \in \Xi,\ t = 1, \ldots, T \end{array} \tag{$\mathcal{SP}$}$$

Robust optimization problems If the distribution governing the uncertainty ξ is unknown or if the decision maker is very risk-averse, then it is not possible or unreasonable to minimize expected costs. In these situations, a decision maker may want to minimize the worst-case costs, where the worst case (maximum) is evaluated with respect to all possible scenarios $\xi \in \Xi$; see, e.g., [40] for a formal justification. Traditionally, such worst-case (robust) optimization problems only involve here-and-now decisions [7, 20]. Problem $\mathcal{P}$ allows us to formulate a multi-stage generalization of robust optimization problems as follows.

$$\begin{array}{ll} \text{minimize} & \displaystyle\max_{\xi \in \Xi} \left(\displaystyle\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t) \right) \\ \text{subject to} & \left. \begin{array}{r} \displaystyle\sum_{s=1}^{t} A_{ts} x_s(\xi^s) \geq b_t(\xi^t) \\ x_t(\xi^t) \geq 0 \end{array} \right\} \forall \xi \in \Xi,\ t = 1, \ldots, T \end{array} \tag{$\mathcal{RO}$}$$

In order to see that $\mathcal{RO}$ is a special case of $\mathcal{P}$, we consider an *epigraph reformulation* of the worst-case objective,

$$\begin{aligned}\max_{\xi\in\Xi}\left(\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)\right) &= \min_{\tau\in\mathbb{R}}\left\{\tau \;:\; \max_{\xi\in\Xi}\left(\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)\right)\leq\tau\right\}\\ &= \min_{\tau\in\mathbb{R}}\left\{\tau \;:\; \sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)\leq\tau \quad \forall\xi\in\Xi\right\},\end{aligned} \tag{2.1}$$

where $\tau \in \mathbb{R}$ represents an auxiliary (deterministic) decision variable. Replacing the worst-case objective in $\mathcal{RO}$ with (2.1) transforms the robust optimization problem $\mathcal{RO}$ into a variant of the stochastic programming problem $\mathcal{SP}$ with a particularly simple objective function (given by τ). As $\mathcal{RO}$ is a special case of $\mathcal{SP}$ and $\mathcal{SP}$ is a special case of $\mathcal{P}$, we conclude that $\mathcal{RO}$ is indeed a special case of $\mathcal{P}$.

Chance-constrained programs Let $\mathbb{P}_\xi$ be the distribution of ξ. We can then formulate a multi-stage generalization of chance-constrained programs as

$$\begin{aligned}&\text{minimize} && \mathbb{E}_\xi\left(\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)\right)\\ &\text{subject to} && \mathbb{P}_\xi\left(\sum_{t=1}^{T} a_{it}^\top x_t(\xi^t)\geq b_i(\xi)\right)\geq 1-\epsilon_i \quad \forall i=1,\ldots,I,\\ &&& x_t(\xi^t)\geq 0 \qquad \forall\xi\in\Xi,\ t=1,\ldots,T,\end{aligned} \tag{$\mathcal{CC}$}$$

where $a_{it}\in\mathbb{R}^{n_t}$, $b_i(\xi)\in\mathbb{R}$ and $\epsilon_i\in(0,1]$. Here, the ith constraint requires that the inequality $\sum_{t=1}^{T} a_{it}^\top x_t(\xi^t)\geq b_i(\xi)$ should be satisfied with probability at least $1-\epsilon_i$. Chance constraints of this type are useful for modeling risk preferences and safety constraints in engineering applications. Note that a chance constraint with $\epsilon_i = 1$ reduces to a robust constraint that must hold for all $\xi\in\Xi$. Therefore, chance constraints with $\epsilon_i < 1$ can be viewed as *soft* versions of the corresponding robust constraints.

We now demonstrate that $\mathcal{CC}$ has a tight conservative approximation of the form $\mathcal{P}$. To this end, we introduce the loss functions $L_i(\xi) = b_i(\xi) - \sum_{t=1}^{T} a_{it}^\top x_t(\xi^t)$. The ith chance constraint is therefore equivalent to the requirement that the smallest $(1-\epsilon_i)$-quantile of the loss distribution, which we denote by $\mathrm{VaR}_{\epsilon_i}(L_i(\xi))$, is non-positive. To obtain a conservative approximation for the chance constraint, we introduce the conditional value-at-risk (CVaR) of $L_i(\xi)$ at level ϵ_i, which is defined as $\mathrm{CVaR}_{\epsilon_i}(L_i(\xi)) = \min_{\beta_i}\{\beta_i + \frac{1}{\epsilon_i}\mathbb{E}_\xi([L_i(\xi) - \beta_i]^+)\}$ [64]. Due to its favorable theoretical and computational properties, CVaR has become a popular risk measure in finance. Rockafellar and Uryasev [68] have shown that the optimal β_i which solves the minimization problem in the definition of CVaR coincides with $\mathrm{VaR}_{\epsilon_i}(L_i(\xi))$ and that the CVaR at level ϵ_i coincides with the conditional expectation of the right tail of the loss distribution above $\mathrm{VaR}_{\epsilon_i}(L_i(\xi))$. Thus, the following implication holds; see also Figure 1.

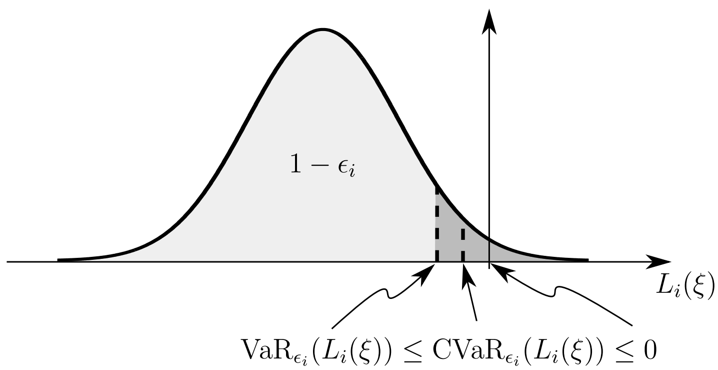

Figure 1. Relationship between $\mathrm{VaR}_{\epsilon_i}(L_i(\xi))$ and $\mathrm{CVaR}_{\epsilon_i}(L_i(\xi))$ for each constraint i, at level ϵ_i.

$$\mathrm{CVaR}_{\epsilon_i}(L_i(\xi)) \leq 0 \quad \Longrightarrow \quad \mathrm{VaR}_{\epsilon_i}(L_i(\xi)) \leq 0 \quad \Longleftrightarrow \quad \mathbb{P}_\xi(L_i(\xi) \leq 0) \geq 1 - \epsilon_i$$

As pointed out by Nemirovski and Shapiro [62], the CVaR constraint on the left-hand side represents the tightest convex approximation for the chance constraint on the right-hand side of the above expression. By linearizing the term $[L_i(\xi) - \beta_i]^+$ in the definition of CVaR, the ith CVaR constraint can be re-expressed as the following system of linear inequalities

$$\beta_i + \frac{1}{\epsilon_i}\mathbb{E}_\xi(z_i(\xi)) \leq 0, \quad z_i(\xi) \geq b_i(\xi) - \sum_{t=1}^{T} a_{it}^\top x_t(\xi^t) - \beta_i, \quad z_i(\xi) \geq 0, \tag{2.2}$$

which involve the deterministic (first stage) variable $\beta_i \in \mathbb{R}$ and a new stochastic (stage-T) variable $z_i(\xi) \in \mathbb{R}$. Replacing each chance constraint in $\mathcal{CC}$ with the corresponding system (2.2) of linear inequalities thus results in a problem of type $\mathcal{P}$ with expectation constraints. Therefore, chance-constrained problems of the type $\mathcal{CC}$ have tight conservative approximations within the class of problems $\mathcal{P}$.

We remark that the general decision problem $\mathcal{P}$ is flexible enough to cover also hybrid models which combine various aspects of deterministic, stochastic, robust and chance-constrained programs in the same model.

3 The Decision Rule Approach

In this section, we derive a tractable approximation to the decision problem $\mathcal{P}$ by restricting the space of the decision rules $x_t(\cdot)$, $t = 1, \ldots, T$, to those that exhibit a linear dependence on the observed problem parameters ξ^t. The second part of the section explains how we can efficiently measure the optimality gap that we incur through this simplification.

3.1 Determining the Best Linear Decision Rule

Problem $\mathcal{P}$ generalizes a number of difficult optimization problems, including multi-stage stochastic programs. It is therefore clear that problem $\mathcal{P}$ is severely computationally intractable itself. A simple but effective approach to improve the tractability of problem $\mathcal{P}$ is to restrict the space of the decision rules $x_t(\cdot)$, $t = 1, \ldots, T$, to those that exhibit a linear dependence on the observation history ξ^t. Remember that we stipulated in Remark 2.1 that $k_1 = 1$ and $\xi_1 = 1$ for all $\xi \in \Xi$. This implies that we actually optimize over all *affine* (*i.e.*, linear plus a constant) decision functions of the *non*-degenerate uncertain parameters $\xi_2, \ldots, \xi_T$ if we optimize over all *linear* functions of $\xi = (\xi_1, \ldots, \xi_T)$.

In the rest of the paper we will assume that the conditional expectations $\mathbb{E}_\xi(\xi|\xi^t)$ are linear in the sense that there exist matrices $M_t \in \mathbb{R}^{k \times k^t}$ such that $\mathbb{E}_\xi(\xi|\xi^t) = M_t \xi^t$ for all $\xi \in \Xi$. This assumption is automatically satisfied, for instance, if the random parameters ξ_t are mutually independent. In this case the conditional expectations reduce to simpler unconditional expectations, and thus we find $\mathbb{E}_\xi(\xi|\xi^t) = (\xi_1, \ldots, \xi_t, \mu_{t+1}, \ldots, \mu_T)$, where μ_t denotes the unconditional mean value of ξ_t. As $\xi_1 = 1$ for all $\xi \in \Xi$, we thus have

$$\mathbb{E}_\xi(\xi|\xi^t) = (\xi_1, \ldots, \xi_t, \mu_{t+1}\xi_1, \ldots, \mu_T \xi_1) \quad \forall \xi \in \Xi.$$

The last expression is manifestly linear in ξ^t. It is easy to verify that the conditional expectations remain linear when the process of the random parameters ξ_t belongs to the large class of autoregressive moving-average models.

For the further argumentation, we define the truncation matrix $P_t \in \mathbb{R}^{k^t \times k}$ through $P_t \xi = \xi^t$. Thus, P_t maps any scenario ξ to the corresponding observation history ξ^t up to stage t. If we model the decision rule $x_t(\xi^t)$ as a linear function of ξ^t, it can thus be expressed as $x_t(\xi^t) = X_t \xi^t = X_t P_t \xi$ for some matrix $X_t \in \mathbb{R}^{n_t \times k^t}$. Substituting these *linear decision rules* into $\mathcal{P}$ yields the following approximate problem.

$$
\begin{array}{ll}
\text{minimize} & \mathbb{E}_\xi \left(\displaystyle\sum_{t=1}^{T} c_t(\xi^t)^\top X_t P_t \xi \right) \\
\text{subject to} & \left. \begin{array}{r} \mathbb{E}_\xi \left(\displaystyle\sum_{s=1}^{T} A_{ts} X_s P_s \xi \;\middle|\; \xi^t \right) \geq b_t(\xi^t) \\ X_t P_t \xi \geq 0 \end{array} \right\} \; \forall \xi \in \Xi, \; t = 1, \ldots, T
\end{array}
\qquad (\mathcal{P}^u)
$$

The objective function of $\mathcal{P}^u$ can be simplified and re-expressed in terms of the second order moment matrix $M = \mathbb{E}(\xi \, \xi^\top)$ of the random parameters, which is not to be confused with the first order conditional moment sensitivity matrices M_t that satisfy $\mathbb{E}_\xi(\xi | \xi^t) = M_t \xi^t$. Interchanging summation and expectation and using the cyclicity property of the trace operator, we obtain

$$
\begin{aligned}
\mathbb{E}_\xi \left(\sum_{t=1}^{T} c_t(\xi^t)^\top X_t P_t \xi \right) &= \sum_{t=1}^{T} \mathbb{E}_\xi(\xi^\top P_t^\top C_t^\top X_t P_t \xi) \\
&= \sum_{t=1}^{T} \mathbb{E}_\xi(\mathrm{Tr}[P_t \xi \xi^\top P_t^\top C_t^\top X_t]) \\
&= \sum_{t=1}^{T} \mathrm{Tr}(P_t M P_t^\top C_t^\top X_t).
\end{aligned}
$$

Similarly, we can reformulate the conditional expectation terms in the constraints of $\mathcal{P}^u$ as

$$
\mathbb{E}_\xi \left(\sum_{s=1}^{T} A_{ts} X_s P_s \xi \;\middle|\; \xi^t \right) = \sum_{s=1}^{T} A_{ts} X_s P_s M_t P_t \xi \,.
$$

Thus, the linear decision rule problem $\mathcal{P}^u$ is equivalent to

$$
\begin{array}{ll}
\text{minimize} & \displaystyle\sum_{t=1}^{T} \mathrm{Tr}(P_t M P_t^\top C_t^\top X_t) \\
\text{subject to} & \left.\begin{array}{r} \left(\displaystyle\sum_{s=1}^{T} A_{ts} X_s P_s M_t P_t - B_t P_t\right) \xi \geq 0 \\ X_t P_t \xi \geq 0 \end{array}\right\} \; \forall \xi \in \Xi, \; t = 1, \ldots, T.
\end{array}
\tag{3.3}
$$

Although problem (3.3) has only finitely many decision variables, that is, the coefficients of the matrices $X_1, \ldots, X_T$ encoding the linear decision rules, it is still not suitable for numerical solution as it involves infinitely many constraints parameterized by $\xi \in \Xi$. The following proposition, which captures the essence of robust optimization, provides the tools for reformulating the ξ-dependent constraints in (3.3) in terms of a finite number of linear constraints [7, 20].

Proposition 3.1. *For any $p \in \mathbb{N}$ and $Z \in \mathbb{R}^{p \times k}$, the following statements are equivalent.*

(i) $Z\xi \geq 0$ *for all* $\xi \in \Xi = \{\xi \in \mathbb{R}^k : W\xi \geq h\}$,

(ii) $\exists \, \Lambda \in \mathbb{R}^{p \times l}$ *with* $\Lambda \geq 0, \; \Lambda W = Z, \; \Lambda h \geq 0.$

Proof. We denote by $Z_\pi^\top$ the πth row of the matrix Z. Then, statement (i) is equivalent to

$$
\left.\begin{array}{lll}
 & Z\xi \geq 0 \text{ for all } \xi \text{ subject to } W\xi \geq h & \\
\iff & 0 \leq \min\limits_{\xi} \left\{ Z_\pi^\top \xi : W\xi \geq h \right\} & \forall \, \pi = 1, \ldots, p \\
\iff & 0 \leq \max\limits_{\Lambda_\pi} \left\{ h^\top \Lambda_\pi : W^\top \Lambda_\pi = Z_\pi, \, \Lambda_\pi \geq 0 \right\} & \forall \, \pi = 1, \ldots, p \\
\iff & \exists \, \Lambda_\pi \text{ with } W^\top \Lambda_\pi = Z_\pi, \, h^\top \Lambda_\pi \geq 0, \, \Lambda_\pi \geq 0 & \forall \, \pi = 1, \ldots, p
\end{array}\right\}
\tag{3.4}
$$

The equivalence in the third line follows from linear programming duality. Interpreting $\Lambda_\pi^\top$ as the πth row of a new matrix $\Lambda \in \mathbb{R}^{p \times l}$ shows that the last line in (3.4) is equivalent to assertion (ii). Thus, the claim follows. □

Using Proposition 3.1, one can reformulate the inequality constraints in (3.3) to obtain

$$\begin{array}{lll} \text{minimize} & \displaystyle\sum_{t=1}^{T} \mathrm{Tr}(P_t M P_t^\top C_t^\top X_t) & \\ \text{subject to} & \left.\begin{array}{r} \displaystyle\sum_{s=1}^{T} A_{ts} X_s P_s M_t P_t - B_t P_t = \Lambda_t W, \Lambda_t h \geq 0, \Lambda_t \geq 0 \\ X_t P_t \; = \Gamma_t W, \Gamma_t h \geq 0, \Gamma_t \geq 0 \end{array}\right\} & \forall t = 1, \ldots, T. \end{array} \qquad (\tilde{\mathcal{P}}^u)$$

The decision variables in $\tilde{\mathcal{P}}^u$ are the entries of the matrices $X_t \in \mathbb{R}^{n_t \times k^t}$, $\Lambda_t \in \mathbb{R}^{m_t \times l}$ and $\Gamma_t \in \mathbb{R}^{n_t \times l}$ for $t = 1, \ldots, T$. Note that the objective function as well as all constraints are linear in these decision variables. Thus, $\tilde{\mathcal{P}}^u$ constitutes a finite linear program, which can be solved efficiently with off-the-shelf solvers such as IBM ILOG CPLEX [1].

A major benefit of using linear decision rules is that the size of the approximating linear program $\tilde{\mathcal{P}}^u$ grows only moderately with the number of time stages. Indeed, the number of variables and constraints is quadratic in k, l, $m = \sum_{t=1}^{T} m_t$, and $n = \sum_{t=1}^{T} n_t$. Note that these numbers usually scale linearly with T, and hence the size of $\tilde{\mathcal{P}}^u$ typically grows only quadratically with the number of decision stages.

We close this section with two remarks about alternative approximation methods to convert problem $\mathcal{P}$ to a finite linear program that is amenable to numerical solution.

Remark 3.2 (Scenario tree approximation). *Instead of approximating the functional form of the decision rules $x_t(\cdot)$, $t = 1, \ldots, T$, we can can improve the tractability of problem $\mathcal{P}$ by replacing the underlying process $\xi_1, \ldots, \xi_T$ of random parameters with a discrete stochastic process. The resulting process can be visualized as a scenario tree, which ramifies at all time points at which new problem data is observed (Figure 2). Scenario tree approaches to stochastic programming have been studied extensively over the past decades; see e.g. the survey paper [32] that accounts for the developments up to the year 2000. More recent contributions are listed in the official stochastic programming bibliography [77]. In contrast to the decision rule approach, scenario tree methods typically scale exponentially with the number of decision stages. Figure 3 compares the scenario tree and the decision rule approximations.*

Remark 3.3 (Sample-Based Optimization). *We derived a tractable approximation for problem $\mathcal{P}$ in two steps. First, we restricted the decision rules $x_t(\cdot)$, $t = 1, \ldots, T$, to be linear functions of the*

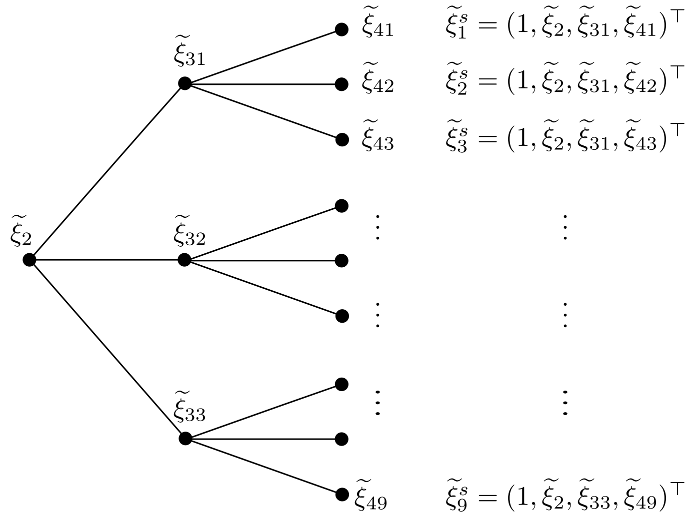

Figure 2. Scenario tree with samples $\widetilde{\xi}^s_1, \ldots, \widetilde{\xi}^s_9$.

observation histories ξ^t. Afterwards, we used linear programming duality to obtain a finite problem. We can derive a different approximation for problem $\mathcal{P}$ if we enforce the semi-infinite constraints in $\mathcal{P}^u$ only over a finite subset of samples $\{\widetilde{\xi}^s_1, \ldots, \widetilde{\xi}^s_K\} \subset \Xi$. It has been shown in [23, 79] that a modest number K of samples suffices to satisfy the semi-infinite constraints in $\mathcal{P}^u$ with high probability. The advantage of such sampling-based approaches is that they allow us to model more general dependencies between the problem data (A, b, c) and the random parameters. However, we are not aware of any methods to measure the optimality gap that we incur with sample-based methods.

3.2 Suboptimality of the Best Linear Decision Rule

The price that we have to pay for the favorable scaling properties of the linear decision rule approximation is a potential loss of optimality. Indeed, the best linear decision rule can result in a substantially higher objective value than the best general decision rule (which is typically non-linear). The difference $\Delta^u = \min \tilde{\mathcal{P}}^u - \min \mathcal{P}$ between the optimal values of the approximate and the original decision problem can be interpreted as the approximation error associated with the linear decision rule approximation. As $\tilde{\mathcal{P}}^u$ is a restriction of the minimization problem $\mathcal{P}$, Δ^u is necessarily non-negative. Modelers should estimate Δ^u in order to assess the appropriateness of the linear decision rule approximation for a particular problem instance: a small Δ^u indicates that implementing the solution of $\tilde{\mathcal{P}}^u$ will incur a negligible loss of optimality, while a large Δ^u may

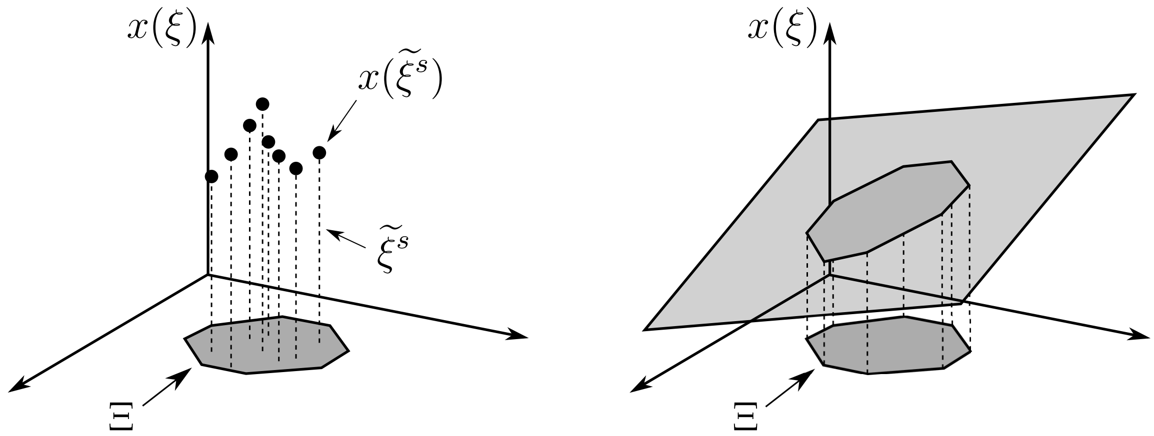

Figure 3. Comparison of the scenario tree (left) and the decision rule approximation (right). Scenario trees replace the process $\xi_1, \ldots, \xi_T$ of random parameters with a discrete stochastic process. The decision rule approach retains the original stochastic process, but it restricts the functional form of the decision rules $x_t(\cdot)$, $t = 1, \ldots, T$.

prompt us to be more cautious and to try to improve the approximation quality (e.g. by using more flexible piecewise linear decision rules; see Section 4).

Generally speaking, there are two ways to measure the approximation error Δ^u. We can derive generic *a priori* bounds on the maximum value of Δ^u that can be incurred over a class of instances, or we can measure Δ^u *a posteriori* for a specific problem instance.

A priori bounds on Δ^u have a long history. In particular, linear decision rules have been proven to optimally solve the linear quadratic regulator problem [9], while piecewise linear decision rules optimally solve two-stage stochastic programs [36]. More recently, linear decision rules have been shown to optimally solve a class of one-dimensional robust control problems [18] and two-stage robust vehicle routing problems [43]. On the other hand, the worst-case approximation ratio for linear decision rules applied to two-stage robust optimization problems with m linear constraints has been shown to be of the order $\mathcal{O}(\sqrt{m})$, see [16]. Similar results have been derived for two-stage stochastic programs in [15].

Given their scarcity and their somewhat limited scope, it seems fair to say that a priori bounds on Δ^u are at most indicative of the expressive power of linear and piecewise linear decision rules. It thus seems natural to consider a posteriori bounds on Δ^u that exploit the specific structure of individual instances of the problem $\mathcal{P}$. Unfortunately, the direct computation of Δ^u for a specific instance of $\mathcal{P}$ would require the solution of $\mathcal{P}$ itself, which is intractable. In this section we demonstrate, however, that an upper bound on Δ^u can be obtained efficiently by studying a dual decision problem associated with $\mathcal{P}$.

It is well known that any primal linear program $\min_x\{c^\top x \;:\; Ax \geq b,\; x \geq 0\}$ has an associated dual linear program $\max_y\{b^\top y \;:\; A^\top y \leq c,\; y \geq 0\}$, which is based on the same problem data (A, b, c), such that the following hold: the minimum of the primal is never smaller than the maximum of the dual (weak duality), and if either the primal or the dual is feasible then the minimum of the primal coincides with the maximum of the dual (strong duality) [29]. There is a duality theory for decision problems of the type $\mathcal{P}$ which is strikingly reminiscent of the duality theory for ordinary linear programs. Following Eisner and Olsen [33], the dual problem corresponding to $\mathcal{P}$ can be defined as

$$\begin{array}{ll} \text{maximize} & \mathbb{E}_\xi\left(\sum_{t=1}^{T} b_t(\xi^t)^\top y_t(\xi^t)\right) \\ \text{subject to} & \left.\begin{array}{r} \mathbb{E}_\xi\left(\sum_{s=1}^{T} A_{st}^\top y_s(\xi^s) \;\middle|\; \xi^t\right) \leq c_t(\xi^t) \\ y_t(\xi^t) \geq 0 \end{array}\right\} \forall \xi \in \Xi,\; t = 1, \dots, T. \end{array} \tag{$\mathcal{D}$}$$

Note that the dual maximization problem $\mathcal{D}$ is stated in terms of the same problem data as the primal minimization problem $\mathcal{P}$. As for ordinary linear programs, dualization transposes the constraint matrices and swaps the roles of the objective function and right-hand side coefficients. Dualization also reverts the temporal coupling of the decision stages in the sense that the sums in the constraints of $\mathcal{D}$ now run over the first index of the constraint matrices. Thus, even if the conditional expectations are redundant in the primal problem $\mathcal{P}$ (that is, if $A_{ts} = 0$ for all $s > t$), the conditional expectations are relevant in the dual problem $\mathcal{D}$ (because typically $A_{ts} \neq 0$ for $s \leq t$). Therefore, in hindsight we realize that the inclusion of conditional expectation constraints in $\mathcal{P}$ was necessary to preserve the symmetry of the employed duality scheme.

As in the case of ordinary linear programming, there exist weak and strong duality results for problems $\mathcal{P}$ and $\mathcal{D}$ [33]. In particular, the minimum of $\mathcal{P}$ is never smaller than the maximum of $\mathcal{D}$ (weak duality), and if some technical regularity conditions hold, then the minimum of $\mathcal{P}$ coincides with the maximum of $\mathcal{D}$ (strong duality).

The symmetry between $\mathcal{P}$ and $\mathcal{D}$ enables us to solve $\mathcal{D}$ with the linear decision rule approach that was originally designed for $\mathcal{P}$. Indeed, if we model the dual decision rule $y_t(\xi^t)$ as a linear function of ξ^t, then it can be expressed as $y_t(\xi^t) = Y_t \xi^t$ for some matrix $Y_t \in \mathbb{R}^{m_t \times k^t}$. Substituting these *dual* linear decision rules into $\mathcal{D}$ yields an approximate problem $\mathcal{D}^l$, which can be shown to be equivalent to the following tractable linear program.

$$\begin{aligned} &\text{maximize} \quad \sum_{t=1}^{T} \text{Tr}(P_t M P_t^\top B_t^\top Y_t) \\ &\text{subject to} \quad \left. \begin{aligned} \sum_{s=1}^{T} A_{st}^\top Y_s P_s M_t P_t - C_t P_t &= \Phi_t W, \Phi_t h \le 0, \Phi_t \le 0 \\ Y_t P_t &= \Psi_t W, \Psi_t h \ge 0, \Psi_t \ge 0 \end{aligned} \right\} \forall t = 1, \ldots, T. \end{aligned} \qquad (\tilde{\mathcal{D}}^l)$$

The decision variables in $\tilde{\mathcal{D}}^l$ are the entries of the matrices $Y_t \in \mathbb{R}^{m_t \times k^t}$, $\Phi_t \in \mathbb{R}^{n_t \times l}$ and $\Psi_t \in \mathbb{R}^{m_t \times l}$ for $t = 1, \ldots, T$.

In analogy to the primal approximation error Δ^u, the dual approximation error can be defined as $\Delta^l = \max \mathcal{D} - \max \tilde{\mathcal{D}}^l$. As $\tilde{\mathcal{D}}^l$ is a restriction of the maximization problem $\mathcal{D}$, Δ^l is necessarily non-negative. It quantifies the loss of optimality of the best *linear* dual decision rule with respect to the best *general* dual decision rule. Unfortunately, Δ^l is usually unknown as its computation would require the solution of the original dual decision problem $\mathcal{D}$. However, the *joint* primal and dual approximation error $\Delta = \min \tilde{\mathcal{P}}^u - \max \tilde{\mathcal{D}}^l$ is efficiently computable; it merely requires the solution of two tractable finite linear programs. Note that Δ constitutes indeed an upper bound on both Δ^u and Δ^l since

$$\begin{aligned} \Delta &= \min \tilde{\mathcal{P}}^u - \max \tilde{\mathcal{D}}^l \\ &= \min \tilde{\mathcal{P}}^u - \min \mathcal{P} + \min \mathcal{P} - \max \mathcal{D} + \max \mathcal{D} - \max \tilde{\mathcal{D}}^l \\ &= \Delta^u + \min \mathcal{P} - \max \mathcal{D} + \Delta^l \\ &\ge \Delta^u + \Delta^l, \end{aligned}$$

where the last inequality follows from weak duality.

We conclude that for any decision problem of the type $\mathcal{P}$ the best primal and dual linear decision rules can be computed efficiently by solving tractable finite linear programs. The corresponding approximation errors are bounded by Δ, which can also be computed efficiently. Moreover, the optimal values of the approximate problems $\tilde{\mathcal{P}}^u$ and $\tilde{\mathcal{D}}^l$ provide upper and lower bounds on the optimal value of the original problem $\mathcal{P}$, respectively.

4 Non-linear Decision Rules

A large value of Δ indicates that either the primal or the dual approximation (or both) are inadequate. If the loss of optimality of linear decision rules is unacceptably high, modelers will endeavor to find a less conservative (but typically more computationally demanding) approximation. Ideally, one would choose a richer class of decision rules over which to optimize.

In this section we show that the techniques developed for *linear* decision rules can also be used to optimize efficiently over more flexible classes of *non-linear* decision rules. The underlying theory has been developed in a series of recent publications [7, 27, 39, 41]. We will motivate the general approach first through an example.

Example 4.1. *Assume that a two-dimensional random vector $\xi = (\xi_1, \xi_2)$ is uniformly distributed on $\Xi = \{1\} \times [-1, 1]$. This choice of Ξ satisfies the standard assumption that $\xi_1 = 1$ for all $\xi \in \Xi$. Any scalar linear decision rule is thus representable as $x(\xi) = X_1\xi_1 + X_2\xi_2$, where X_1 denotes a constant offset (since ξ_1 is equal to 1 with certainty), while X_2 characterizes the sensitivity of the decision with respect to ξ_2; see Figure 4 (left). To improve flexibility, one may introduce a breakpoint at $\xi_2 = 0$ and consider piecewise linear continuous decision rules that are linear in ξ_2 on the subintervals $[-1, 0]$ and $[0, 1]$, respectively. These decision rules are representable as*

$$x(\xi) = X_1'\xi_1 + X_2'(\min\{\xi_2, 0\}) + X_3'(\max\{\xi_2, 0\}), \tag{4.5}$$

where X_1' denotes again a constant offset, while X_2' and X_3' characterize the sensitivities of the decision with respect to ξ_2 on the subintervals $[-1, 0]$ and $[0, 1]$, respectively; see Figure 4 (center). We can now define a new set of random variables $\xi_1' = \xi_1$, $\xi_2' = \min\{\xi_2, 0\}$ and $\xi_3' = \max\{\xi_2, 0\}$, which are completely determined by ξ. In particular, the support for $\xi' = (\xi_1', \xi_2', \xi_3')$ is given by $\Xi' = \{\xi' \in \mathbb{R}^3 : \xi_1' = 1,\ \xi_2' \in [-1, 0],\ \xi_3' \in [0, 1],\ \xi_2'\xi_3' = 0\}$. Note also that ξ' is uniformly distributed on Ξ'. We will henceforth refer to ξ' as the lifted random vector *as it ranges over a higher-dimensional 'lifted' space. Moreover, the function*

$$L(\xi) = (L_1(\xi), L_2(\xi), L_3(\xi)) = (\xi_1, \min\{\xi_2, 0\}, \max\{\xi_2, 0\}),$$

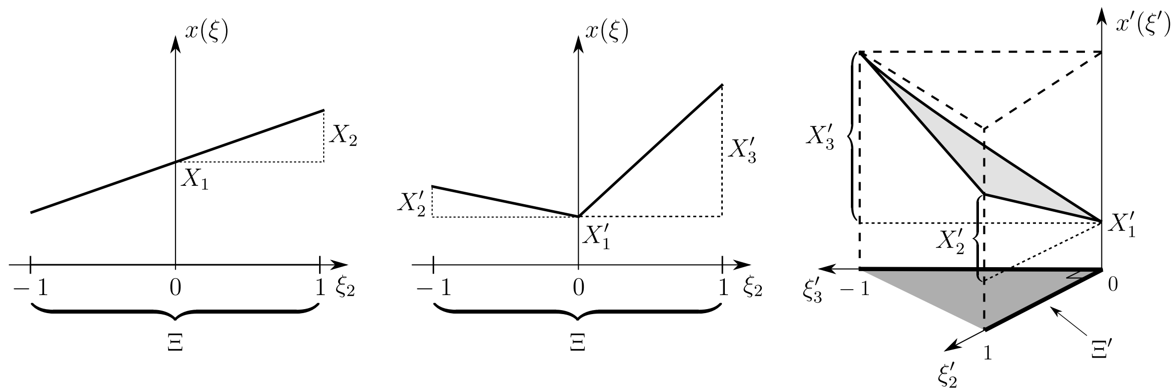

Figure 4. Illustration of the linear and piecewise linear decision rules in the original and the lifted space. Note that the set $\Xi' = L(\Xi)$ is non-convex, represented by the thick line in the right diagram. Since the decision rule $x'(\xi')$ is a linear function of the random parameters ξ', however, it is non-negative over Ξ' if and only if it is non-negative over the convex hull of Ξ', which is given by the dark shaded region.

which maps ξ to ξ', will be referred to as a lifting*. By construction, the piecewise linear decision rule* (4.5) *in the original space is equivalent to the linear decision rule $x'(\xi') = X_1'\xi_1' + X_2'\xi_2' + X_3'\xi_3'$ in the lifted space; see Figure 4 (right). Moreover, due to the linearity of $x'(\xi')$ in ξ', the decision rule $x'(\xi')$ is non-negative over the non-convex set Ξ' if and only if $x'(\xi')$ is non-negative over the convex hull of Ξ'. Similarly, $x'(\xi')$ satisfies an arbitrary linear inequality (with constant coefficients and with intercepts that are linear in ξ') uniformly over the non-convex set Ξ' if and only if $x'(\xi')$ satisfies the same inequality uniformly over the convex hull of Ξ'. We can therefore replace the non-convex support Ξ' in the lifted space with its (polyhedral) convex hull, which can be represented as an intersection of halfspaces as required in Section 2.1. Hence, all techniques developed for linear decision rules can also be used for piecewise linear decision rules of the form* (4.5)*. Recipes for constructing more flexible piecewise linear decision rules are reported in [39, Section 4] and [41, Section 4]. More general non-linear decision rules are discussed in [39, Section 5].*

To solve a general decision problem of the type $\mathcal{P}$ in non-linear decision rules, we define a *lifting operator*

$$L(\xi) = (L_1(\xi_1), \ldots, L_T(\xi_T)),$$

where each $L_t(\xi_t)$ represents a continuous function from $\mathbb{R}^{k_t}$ to $\mathbb{R}^{k_t'}$ for some $k_t' \geq k_t$. Using the lifting operator, we can construct a lifted random vector $\xi' = (\xi'_1, \ldots, \xi'_T) \in \mathbb{R}^{k'}$, where $\xi'_t = L_t(\xi_t)$, $t = 1, \ldots, T$, and $k' = k'_1 + \cdots + k'_T$. The distribution of the lifted random vector ξ' is completely determined by that of the primitive random vector ξ, and the support of ξ' can be defined as $\Xi' = L(\Xi)$. As for the primitive uncertainties it proves useful to define observation histories $\xi'^t = (\xi'_1, \ldots, \xi'_t) \in \mathbb{R}^{k'^t}$, $k'^t = k'_1 + \cdots + k'_t$, and truncation operators $P'_t : R^{k'} \to \mathbb{R}^{k'^t}$ which map ξ' to ξ'^t, respectively. Our goal is to solve problem $\mathcal{P}$ in non-linear decision rules of the form $x_t(\xi^t) = X'_t P'_t L(\xi)$, which constitute linear combinations of the component functions of the lifting operator. The matrices $X'_t \in \mathbb{R}^{n_t \times k'^t}$ contain the coefficients of these linear combinations, while the truncation operators P'_t eliminate those components of $L(\xi)$ that depend on the future uncertainties $\xi_{t+1}, \ldots, \xi_T$, thereby ensuring non-anticipativity. By construction, the non-linear decision rules $x_t(\xi^t) = X'_t P'_t L(\xi)$ depending on the primitive uncertainties are equivalent to linear decision rules $x'_t(\xi'^t) = X'_t \xi'^t$ depending on the lifted uncertainties. Note that ordinary linear decision rules in the primitive uncertainties can be recovered by choosing a trivial lifting operator $L(\xi) = \xi$.

For the further argumentation we require that the lifting preserves the degeneracy of the first random parameter, that is, $\xi'_1 = 1$ for all $\xi' \in \Xi'$. Moreover, we assume that there is a linear *retraction operator*

$$R(\xi') = (R_1(\xi'_1), \ldots, R_T(\xi'_T)),$$

which allows us to express the primitive random vector ξ as a linear function of the lifted random vector ξ'. To this end, we assume that each R_t represents a linear function from $\mathbb{R}^{k'_t}$ to $\mathbb{R}^{k_t}$. The best non-linear decision rule can then be computed by solving the following optimization problem.

$$\begin{aligned}
&\text{minimize} \quad \mathbb{E}_{\xi'}\left(\sum_{t=1}^{T} c_t(P_t R(\xi'))^\top X'_t P'_t \xi'\right) \\
&\text{subject to} \quad \left.\begin{aligned} \mathbb{E}_{\xi'}\left(\sum_{s=1}^{T} A_{ts} X'_s P'_s \xi' \mid \xi^t\right) \geq b_t(P_t R(\xi')) \\ X'_t P'_t \xi' \geq 0 \end{aligned}\right\} \; \forall \xi' \in \Xi',\ t = 1, \ldots, T
\end{aligned} \qquad (\mathcal{P}'^u)$$

Note that the observation histories in the objective function and in the right-hand side coefficients have been expressed as $\xi^t = P_t \xi = P_t R(\xi')$. The optimization variables in $\mathcal{P}'^u$ are the entries of the matrices $X'_t \in \mathbb{R}^{n_t \times k'^t}$. The key observation is that the approximate problems $\mathcal{P}'^u$ and

$\mathcal{P}^u$ have exactly the same structure. The only difference is that $\Xi' = L(\Xi)$ is typically not a polyhedron because of the non-linearity of the lifting operator. In the following, we assume that an exact representation (or outer approximation) of the convex hull of Ξ' is available in the form of l' inequality constraints:

$$\hat{\Xi} := \{\xi' \in \mathbb{R}^{k'} : W'\xi' \geq h'\},$$

where $\operatorname{conv}\Xi' = \hat{\Xi}$ (exact representation) or $\operatorname{conv}\Xi' \subset \hat{\Xi}$ (outer approximation). Such representations can be determined efficiently for polyhedral supports, see [39]. The non-linear decision rule problem $\mathcal{P}'^u$ can then be transformed into a tractable linear program in the same way as the linear decision rule problem $\mathcal{P}^u$ was converted to $\tilde{\mathcal{P}}^u$, see Section 3.

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{t=1}^{T} \operatorname{Tr}(P_t' M' R^\top P_t^\top C_t^\top X_t') \\ \text{subject to} & \left.\begin{array}{r} \displaystyle\sum_{s=1}^{T} A_{ts} X_s' P_s' M_t' P_t' - B_t P_t R = \Lambda_t W', \Lambda_t h' \geq 0, \Lambda_t \geq 0 \\ X_t' P_t' = \Gamma_t W', \Gamma_t h' \geq 0, \Gamma_t \geq 0 \end{array}\right\} \forall t = 1, \ldots, T \end{array} \qquad (\tilde{\mathcal{P}}'^u)$$

Here, the matrix $R \in \mathbb{R}^{k \times k'}$ is defined through $R\xi' = R(\xi')$ for all $\xi' \in \Xi'$, $M' = \mathbb{E}_{\xi'}(\xi'\xi'^\top) \in \mathbb{R}^{k' \times k'}$ denotes the second order moment matrix associated with ξ', and the conditional expectations are assumed to satisfy $\mathbb{E}_{\xi'}(\xi'|\xi'^t) = M_t'\xi'^t$ for some matrices $M_t' \in \mathbb{R}^{k' \times k'^t}$ and all $\xi' \in \Xi'$. The decision variables in $\tilde{\mathcal{P}}'^u$ are the entries of the matrices $X_t' \in \mathbb{R}^{n_t \times k'^t}$, $\Lambda_t \in \mathbb{R}^{m_t \times l'}$ and $\Gamma_t \in \mathbb{R}^{n_t \times l'}$ for $t = 1, \ldots, T$.

Similarly, we can measure the suboptimality of non-linear decision rules by solving the following dual approximate problem (see Section 3.2).

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{t=1}^{T} \operatorname{Tr}(P_t' M' R P_t^\top B_t^\top Y_t') \\ \text{subject to} & \left.\begin{array}{r} \displaystyle\sum_{s=1}^{T} A_{st}^\top Y_s' P_s' M_t' P_t' - C_t P_t R = \Phi_t W', \Phi_t h' \leq 0, \Phi_t \leq 0 \\ Y_t' P_t = \Psi_t W', \Psi_t h' \geq 0, \Psi_t \geq 0 \end{array}\right\} \forall t = 1, \ldots, T \end{array} \qquad (\tilde{\mathcal{D}}^l)$$

The decision variables of this problem are the entries of the matrices $Y_t' \in \mathbb{R}^{m_t \times k'^t}$, $\Phi_t \in \mathbb{R}^{n_t \times l'}$ and $\Psi_t \in \mathbb{R}^{m_t \times l'}$ for $t = 1, \ldots, T$. One can show that the finite-dimensional primal approximation $\tilde{\mathcal{P}}'^u$

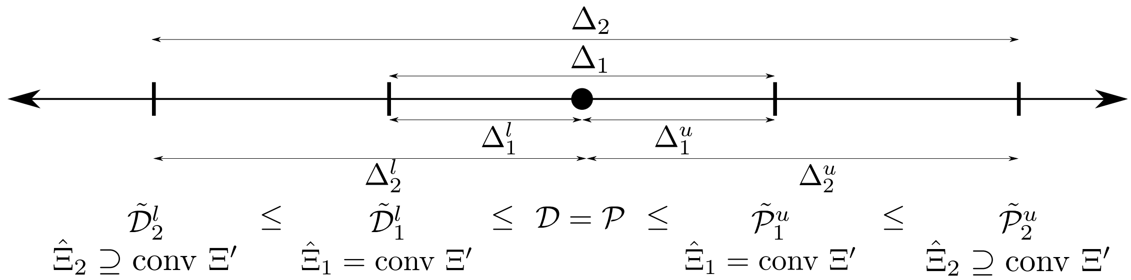

Figure 5. Relationship between the primal bounds $\tilde{\mathcal{P}}_i^u$ and the dual bounds $\tilde{\mathcal{D}}_i^l$ for an exact representation of $\operatorname{conv}\Xi'$ ($i = 1$) and an outer approximation of $\operatorname{conv}\Xi'$ ($i = 2$).

is *equivalent* to the semi-infinite primal problem $\mathcal{P}'^u$ if $\hat{\Xi}$ coincides with $\operatorname{conv}\Xi'$, and $\tilde{\mathcal{P}}'^u$ provides an *upper* bound on the optimal value of $\mathcal{P}'^u$ if $\hat{\Xi}$ is an outer approximation of $\operatorname{conv}\Xi'$. Likewise, the finite-dimensional dual approximation $\tilde{\mathcal{D}}'^l$ is *equivalent* to the semi-infinite dual problem $\mathcal{D}'^l$ (not shown here) if $\hat{\Xi}$ coincides with $\operatorname{conv}\Xi'$, and $\tilde{\mathcal{D}}'^l$ provides a *lower* bound on the optimal value of $\mathcal{D}'^l$ if $\hat{\Xi}$ is an outer approximation of $\operatorname{conv}\Xi'$. In particular, the finite-dimensional approximate problems $\tilde{\mathcal{P}}'^u$ and $\tilde{\mathcal{D}}'^l$ still bracket the optimal value of the problem $\mathcal{P}$ in non-linear decision rules if we employ an outer approximation $\hat{\Xi}$ of the convex hull of Ξ'. The situation is illustrated in Figure 5.

5 Incorporating Integer Decisions

Optimization problems often involve decisions that are modeled through integer variables. These problems are still amenable to the decision rule techniques described in Sections 3 and 4 if the integer variables do not depend on the uncertain problem parameters ξ. In order to substantiate this claim, we consider a variant of problem $\mathcal{P}$ in which the right-hand side vectors of the constraints may depend on a vector of integer variables $z \in \mathbb{Z}^d$.

$$
\begin{array}{ll}
\text{minimize} & \mathbb{E}_\xi\left(\displaystyle\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)\right) \\
\text{subject to} & z \in \mathcal{Z} \\
& \left.\begin{array}{r} \mathbb{E}_\xi\left(\displaystyle\sum_{s=1}^{T} A_{ts} x_s(\xi^s) \;\middle|\; \xi^t\right) \geq b_t(z, \xi^t) \\ x_t(\xi^t) \geq 0 \end{array}\right\} \forall \xi \in \Xi,\ t = 1, \ldots, T
\end{array}
\qquad (\mathcal{P}_{\text{MI}})
$$

As usual, we assume that $b_t(z, \xi^t)$ depends linearly on ξ^t, that is, $b_t(z, \xi^t) = B_t(z)\xi^t$ for some matrix $B_t(z) \in \mathbb{R}^{m_t \times k^t}$. We further assume that $B_t(z)$ depends linearly on z and that $\mathcal{Z} \subset \mathbb{R}^d$ results from the intersection of $\mathbb{Z}^d$ with a convex compact polytope.

In order to apply the decision rule techniques from Sections 3 and 4 to $\mathcal{P}_{\text{MI}}$, we study the parametric program $\mathcal{P}(z)$, which is obtained from $\mathcal{P}_{\text{MI}}$ by fixing the integer variables z. By construction, $\mathcal{P}(z)$ is a decision problem of the type $\mathcal{P}$, which is bounded above and below by the linear programs $\tilde{\mathcal{P}}^u(z)$ and $\tilde{\mathcal{D}}^l(z)$ associated with the primal and dual linear decision rule approximations, respectively. Thus, an upper bound on $\mathcal{P}_{\text{MI}}$ is obtained by minimizing the optimal value of $\tilde{\mathcal{P}}^u(z)$ over all $z \in \mathcal{Z}$. The resulting optimization problem, which we denote by $\tilde{\mathcal{P}}^u_{\text{MI}}$, represents a mixed-integer linear program (MILP).

$$
\begin{array}{ll}
\text{minimize} & \displaystyle\sum_{t=1}^{T} \text{Tr}(P_t M P_t^\top C_t^\top X_t) \\
\text{subject to} & z \in \mathcal{Z} \\
& \left.\begin{array}{r} \displaystyle\sum_{s=1}^{T} A_{ts} X_s P_s M_t P_t - B_t(z) P_t = \Lambda_t W, \Lambda_t h \geq 0, \Lambda_t \geq 0 \\ X_t P_t \;= \Gamma_t W, \Gamma_t h \geq 0, \Gamma_t \geq 0 \end{array}\right\} \forall t = 1, \ldots, T
\end{array}
\qquad (\tilde{\mathcal{P}}^u_{\text{MI}})
$$

Similarly, a lower bound on $\mathcal{P}_{\text{MI}}$ is obtained by minimizing the optimal value of $\tilde{\mathcal{D}}^l(z)$ over all $z \in \mathcal{Z}$. The resulting min-max problem has a bilinear objective function that is linear in the integer variables z and in the coefficients of the dual decision rules Y_t. In order to convert this problem to an MILP, we follow the exposition in [53] and dualize $\tilde{\mathcal{D}}^l(z)$. One can show that strong duality holds whenever the original problem $\mathcal{P}_{\text{MI}}$ is feasible. By construction, we thus obtain a lower bound on $\mathcal{P}_{\text{MI}}$ minimizing the dual of $\tilde{\mathcal{D}}^l(z)$ over all $z \in \mathcal{Z}$. The resulting optimization problem, which we denote by $\tilde{\mathcal{P}}^l_{\text{MI}}$, again represents an MILP.

$$
\begin{array}{ll}
\text{minimize} & \displaystyle\sum_{t=1}^{T} \text{Tr}(P_t M P_t^\top C_t^\top X_t) \\
\text{subject to} & z \in \mathcal{Z} \\
& \left.\begin{array}{r} \displaystyle\sum_{s=1}^{T} A_{ts} X_s P_s M_t P_t + S_t P_t = B_t(z) P_t \\ \left(W - h\, e_1^\top\right) M P_t^\top X_t^\top \geq 0 \\ \left(W - h\, e_1^\top\right) M P_t^\top S_t^\top \geq 0 \end{array}\right\} \forall t = 1, \ldots, T
\end{array}
\tag{$\tilde{\mathcal{P}}^l_{\text{MI}}$}
$$

The optimization variables in $\tilde{\mathcal{P}}^l_{\text{MI}}$ are the matrices $X_t \in \mathbb{R}^{n_t \times k^t}$ and $S_t \in \mathbb{R}^{m_t \times k^t}$, as well as the binary variables $z \in \mathcal{Z}$. Problem $\mathcal{P}_{\text{MI}}$ is also amenable to the refined approximation methods based on piecewise linear decision rules as discussed in Section 4. Further details can be found in [39].

Remark 5.1. *We assumed that the integer decisions z in problem $\mathcal{P}_{\text{MI}}$ do not impact the coefficients $c_t(\cdot)$ and A of the objective function and the constraints, respectively. It is straightforward to apply the techniques presented in this section to a generalized problem where the objective function and constraint coefficients depend linearly on z. Using a Big-M reformulation, the resulting primal and dual approximate problems can again be cast as MILPs.*

6 Case Studies

In the following, we apply the decision rule approach to two well-known operations management problems, and we compare the method with alternative approaches to account for data uncertainty. All of our numerical results are obtained using the IBM ILOG CPLEX 12 optimization package on a dual-core 2.4GHz machine with 4GB RAM [1].

6.1 Production Planning

Our first case study concerns a medium-term production planning problem for a multiproduct plant with uncertain customer demands and backlogging. We assume that the plant consists of a single processing unit that is capable of manufacturing different products in a continuous single-stage process. We first elaborate a formulation that disregards changeovers, and we afterwards extend the model to sequence-dependent changeover times and costs.

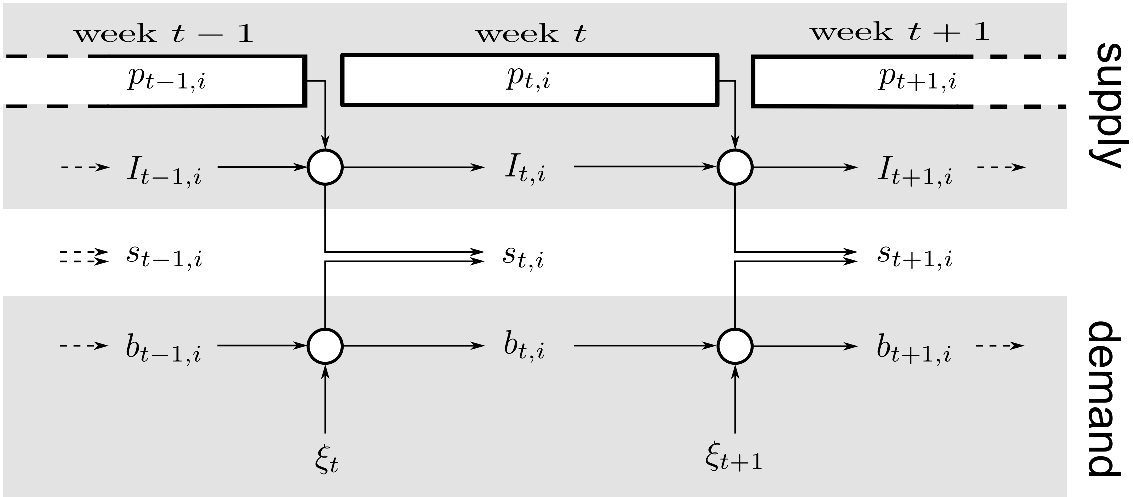

Figure 6. Temporal structure of the production planning model. The decisions with subscript t may depend on all demands realized in weeks $1, \ldots, t$.

We wish to determine a production plan that maximizes the expected profit for a set of products $\mathcal{I}$ and a weekly planning horizon $\mathcal{T} = \{1, \ldots, T\}$. To this end, we denote by p_{ti} the amount of product $i \in \mathcal{I}$ that is produced during week $t \in \mathcal{T} \setminus \{T\}$. The processing unit can manufacture r_i units of product i per hour, and it has an uptime of $\overline{R}$ hours per week. At the beginning of each week $t \in \mathcal{T}$, we observe the demand ξ_{ti} that arises for product i during week t. We assume that the demands ξ_{1i} in the first week are deterministic, while the other demands ξ_{ti}, $t > 1$, are stochastic. Having observed the demands ξ_{ti}, we then decide on the quantity s_{ti} of product i that we sell during week t at a unit price P_{ti}. We also determine the orders b_{ti} for product i that we backlog during week t at a unit cost CB_{ti}. We assume that the sales s_{ti} in week t must be served from the stock produced in week $t-1$ or before. Once the sales decisions s_{ti} for week t have been made, the inventory level I_{ti} for product i during week t is known. Each unit of product i held during period t leads to inventory holding costs CI_{ti}, and we require that the inventory levels I_{ti} satisfy the lower and upper inventory bounds $\underline{I}_{ti}$ and $\overline{I}_{ti}$, respectively. Deterministic versions of this problem have been studied in [26, 56]. For literature surveys on related production planning problems, we refer to [34, 80]. The temporal structure of the problem is illustrated in Figure 6.

We can formulate the production planning problem as follows.

$$
\begin{aligned}
\text{maximize} \quad & \mathbb{E}\left[\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}} P_{ti}s_{ti}(\xi^t) - CB_{ti}b_{ti}(\xi^t) - CI_{ti}I_{ti}(\xi^t)\right] \\
\text{subject to} \quad & \left.\begin{aligned} & \sum_{i\in\mathcal{I}} p_{ti}(\xi^t)/r_i \leq \overline{R} \\ & p_{ti}(\xi^t) \geq 0 \qquad \forall i \in \mathcal{I} \end{aligned}\right\} \forall t \in \mathcal{T}\setminus\{T\} \\
& \left.\begin{aligned} & b_{ti}(\xi^t) = b_{t-1,i}(\xi^{t-1}) + \xi_{ti} - s_{ti}(\xi^t) \\ & I_{ti}(\xi^t) = I_{t-1,i}(\xi^{t-1}) + p_{t-1,i}(\xi^{t-1}) - s_{ti}(\xi^t) \end{aligned}\right\} \forall t \in \mathcal{T}\setminus\{1\},\ \forall i \in \mathcal{I} \\
& b_{1i}(\xi^t) = b_{0i} + \xi_{1i} - s_{1i}(\xi^1), \quad I_{1i}(\xi^t) = I_{0i} - s_{1i}(\xi^1) \\
& b_{ti}(\xi^t) \geq 0, \quad s_{ti}(\xi^t) \geq 0, \quad \underline{I}_{ti} \leq I_{ti}(\xi^t) \leq \overline{I}_{ti} \qquad \forall t \in \mathcal{T},\ \forall i \in \mathcal{I}
\end{aligned}
$$

We require the constraints to be satisfied for all realizations $\xi \in \Xi$ of the uncertain customer demands. The parameters b_{0i} and I_{0i} specify the initial backlog and inventory, respectively. One can easily show that the production planning problem is an instance of the problem $\mathcal{P}$ studied in Section 2.1. The same applies to variations of the problem where the prices and/or costs are uncertain, as well as variations where the product demand is only known at the end of each week. For the sake of brevity, we disregard these variants here.

So far, our production planning problem does not account for changeovers between consecutively manufactured products. Frequent changeovers are undesirable as the involved clean-up, set-up and start-up activities result in both delays and costs. To incorporate changeovers, we follow the approach presented in [56] and introduce binary variables c_{tij}, $t \in \mathcal{T}\setminus\{T\}$ and $i, j \in \mathcal{I}$, that indicate whether a changeover from product i to product j occurs in week t. Likewise, we introduce binary variables c'_{tij}, $t \in \mathcal{T} \setminus \{T\}$ and $i, j \in \mathcal{I}$, that indicate whether a changeover from product i to product j occurs between weeks $t-1$ and t. We set $c_{tii} = 0$, that is, no changeover occurs during the manufacturing of product i, and $c'_{1ij} = 0$ for $i, j \in \mathcal{I}$, that is, no changeover is required for the first product in the first week. Since c_{tij} and c'_{tij} are binary, we must choose the changeovers for all weeks as a here-and-now decision in our production planning model (see Section 5). Note, however, that the actual production amounts p_{ti}, sales decisions s_{ti} and backlogged demands b_{ti} remain wait-and-see decisions that can adapt to the realization of the uncertain customer demands ξ^t.

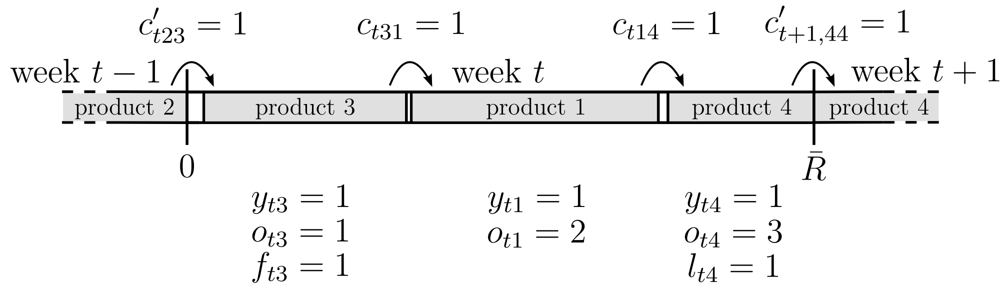

Figure 7. The definition of the changeover variables c_{tij} and c'_{tij} is enforced through the auxiliary variables y_{tj}, o_{tj} and f_{tj}, l_{tj}.

Our interpretation of the changeover variables c_{tij} and c'_{tij} is enforced as follows.

$$\left.\begin{aligned} &\sum_{i\in\mathcal{I}} c_{tij} = y_{tj} - f_{tj} \\ &\sum_{i\in\mathcal{I}} c_{tji} = y_{tj} - l_{tj} \end{aligned}\right\} \forall j \in \mathcal{I},\ \forall t \in \mathcal{T}\setminus\{T\}, \qquad \left.\begin{aligned} &\sum_{i\in\mathcal{I}} c'_{tij} = f_{tj} \\ &\sum_{i\in\mathcal{I}} c'_{tji} = l_{t-1,j} \end{aligned}\right\} \forall j \in \mathcal{I},\ \forall t \in \mathcal{T}\setminus\{1,T\},$$

$$\sum_{i\in\mathcal{I}} f_{ti} = \sum_{i\in\mathcal{I}} l_{ti} = 1 \quad \forall t \in \mathcal{T}\setminus\{T\}, \qquad \left.\begin{aligned} &f_{tj} \le y_{tj} \\ &l_{tj} \le y_{tj} \end{aligned}\right\} \forall j \in \mathcal{I},\ \forall t \in \mathcal{T}\setminus\{T\},$$

$$\left.\begin{aligned} &o_{tj} \le M\, y_{tj} \\ &f_{tj} \le o_{tj} \le \sum_{i\in\mathcal{I}} y_{ti} \\ &o_{tj} \ge o_{ti} + 1 - M(1 - c_{tij}) \quad \forall i \in \mathcal{I} \end{aligned}\right\} \forall j \in \mathcal{I},\ t \in \mathcal{T}\setminus\{T\}.$$

Here, M denotes a sufficiently large constant. The binary variables y_{tj}, f_{tj} and l_{tj} indicate whether product j is manufactured, manufactured first and manufactured last in week t, respectively, and the continuous variables o_{tj} determine the production order in week t. The changeover variables and constraints are illustrated in Figure 7.

To incorporate changeovers in our model, we replace the objective function with

$$\mathbb{E}\left[\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}} P_{ti}s_{ti}(\xi^t) - CB_{ti}b_{ti}(\xi^t) - CI_{ti}I_{ti}(\xi^t)\right] - \sum_{t\in\mathcal{T}\setminus\{T\}}\sum_{i,j\in\mathcal{I}} CC_{ij}\left(c_{tij} + c'_{tij}\right),$$

where CC_{ij} represents the costs of a changeover from product i to product j (with $CC_{ii} = 0$).

	Product 1	Product 2	Product 3	Product 4	Product 5
d_i	10,000	25,000	30,000	30,000	30,000
P_{ti}	0.25	0.40	0.65	0.55	0.45
CB_{ti}	0.05	0.08	0.13	0.11	0.09
r_i	800	900	1,000	1,000	1,200

Table 1. Parameters for the production planning instance. The product prices P_{ti} and backlogging costs CB_{ti} are assumed to be time-invariant.

Likewise, we replace the first constraint of our production planning model with

$$\sum_{i \in \mathcal{I}} p_{ti}(\xi^t)/r_i + \sum_{i,j \in \mathcal{I}} \tau_{ij}(c_{tij} + c'_{tij}) \leq \overline{R} \qquad \forall t \in \mathcal{T} \setminus \{T\},$$

where τ_{ij} denotes the duration of a changeover from product i to product j (with $\tau_{ii} = 0$). One readily verifies that the production planning problem with changeover constraints is an instance of the problem $\mathcal{P}_{\mathrm{MI}}$ studied in Section 5.

6.1.1 Numerical Results

We consider an instance of the production planning problem with 5 products. The nominal demand ϱ_{ti} for product i in week t follows a cyclical pattern:

$$\varrho_{ti} = \left[1 + \frac{1}{2} \sin\left(\frac{\pi(t-2)}{26}\right)\right] d_i \qquad \forall t \in \mathcal{T},\ \forall i \in \mathcal{I},$$

where d_i denotes the long-term average demand for product i. We model the actual demands ξ_{ti} as independent and uniformly distributed random variables with support $[(1-\theta)\varrho_{ti}, (1+\theta)\varrho_{ti}]$, where θ denotes the level of uncertainty. We set the weekly uptime of the processing unit to $\overline{R} = 168$h, and we select inventory holding costs $CI_{ti} = 3.06 \cdot 10^{-5}$ and inventory bounds $(\underline{I}_{ti}, \overline{I}_{ti}) = (0, 10^6)$ that are independent of the indices t and i. There are no initial inventories I_{0i} or backlogs b_{0i}, and the remaining problem parameters are defined in Tables 1 and 2. The time horizon $\mathcal{T}$ and the uncertainty level θ are kept flexible.

Consider the production planning problem without changeover constraints. Figure 8 compares the decision rule approximation with classical scenario-based stochastic programming. The first two graphs report the gaps between the primal and dual objective values if we employ linear and piecewise linear decision rules for various time horizons $\mathcal{T}$. The gap amounts to less than 15% for

(CC_{ij}, τ_{ij})	Product 1	Product 2	Product 3	Product 4	Product 5
Product 1	(0,0.00)	(760,2.00)	(760,1.50)	(750,1.00)	(760,0.75)
Product 2	(745,1.00)	(0,0.00)	(750,2.00)	(770,0.75)	(740,0.50)
Product 3	(770,1.00)	(760,1.25)	(0,0.00)	(765,1.50)	(765,2.00)
Product 4	(740,0.50)	(740,1.00)	(745,2.00)	(0,0.00)	(750,1.75)
Product 5	(740,0.70)	(740,1.75)	(750,2.00)	(750,1.50)	(0,0.00)

Table 2. Parameters for the production planning instance (continued).

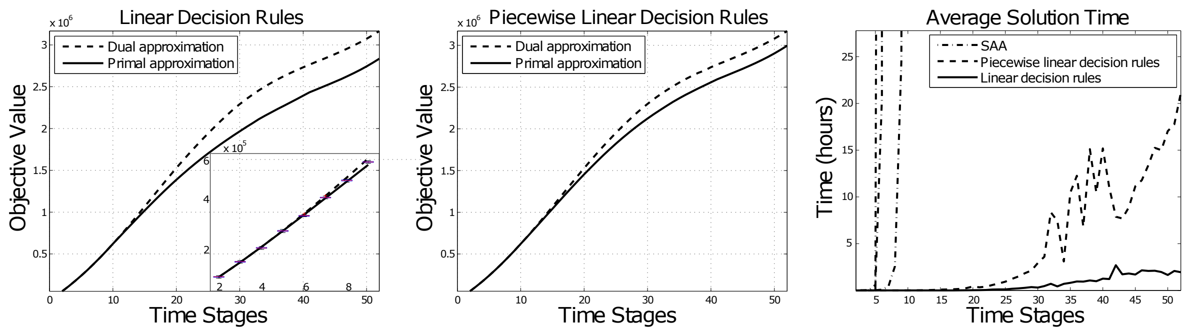

Figure 8. Objective values and runtimes for decision rules and scenario-based stochastic programming. All graphs are based on the uncertainty level $\theta = 0.2$.

linear decision rules, and it can be reduced to about 5% for piecewise linear decision rules with one breakpoint. The first graph also shows a box-and-whisker plot of the objective values reported by the scenario-based formulation for 100 statistically independent scenario trees with a branching factor of two. The third graph illustrates the runtimes required to solve the linear and piecewise linear decision rule problems, as well as the runtimes required by the scenario-based formulation with a branching factor of two, three and four. The figure clearly demonstrates that the employed scenario-based approaches become computationally intractable for problems with many stages and/or a high branching factor. We remark, however, that even a branching factor of four is very small for a problem with five random variables per time stage. Indeed, a branching factor of at least six is required to avoid perfect correlations among the product demands. The problem formulation using decision rules, on the other hand, can be solved for planning horizons of 52 weeks and more.

Consider now the production planning problem with changeover constraints. We want to compare our stochastic problem formulation with a deterministic model that replaces the uncertain customer demands ξ_{ti} with their expected values ϱ_{ti}. To this end, we consider a time horizon of 52 weeks and solve both models in a rolling horizon implementation over a reduced time horizon $\mathcal{T} = \{1, \ldots, 4\}$. At the beginning of each of the 52 weeks, we observe the random demands ξ_{ti}.

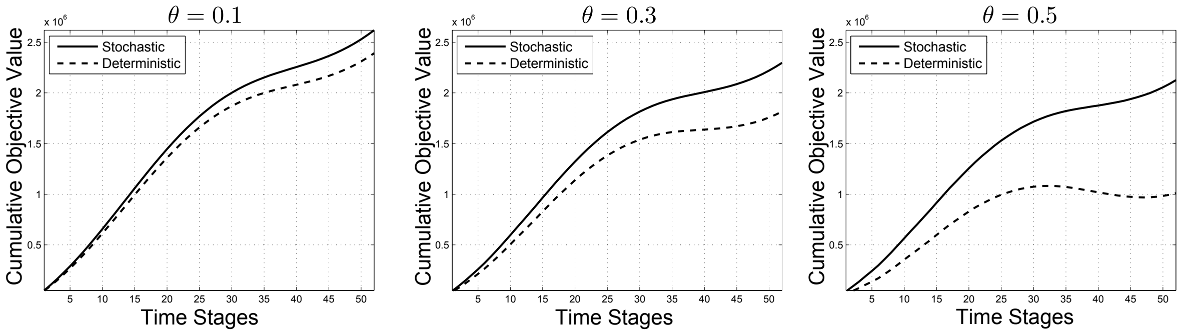

Figure 9. Comparison of piecewise linear decision rules ('stochastic', using one breakpoint) with a deterministic production planning model in a backtest.

We then solve the deterministic and the stochastic model over four weeks. In both models, we set the initial backlogs b_{0i} and the initial stocks I_{0i} to the corresponding values at the end of the previous week, and we adapt the models to account for the initial changeovers c'_{1ij}. We implement the first-stage decisions of both models and repeat the process for the next set of demands $\xi_{t+1,i}$. Figure 9 shows the average cumulative profit achieved by the deterministic and the stochastic model over 100 repetitions of this backtest. The deterministic production planning model performs only slightly worse than the stochastic formulation if the uncertainty level θ is small. For larger values of θ, however, it becomes essential to properly account for the random nature of the customer demands.

6.2 Supply Chain Design

Our second case study investigates the design of multi-echelon supply chains under demand uncertainty. The supply chain produces a single product, and it consists of one production facility and multiple warehouses and distribution centers. The product is manufactured at the production facility, and it is shipped first to the warehouses and then to the distribution centers. Customer demands only arise at the distribution centers, and we do not allow direct deliveries from the production facility to the distribution centers. We assume that the locations of the production facility and the distribution centers are given, whereas there are multiple candidate locations for the warehouses. We denote by $\mathcal{I}$ and $\mathcal{J}$ the index sets of candidate warehouse locations and distribution centers, respectively. We want to build at most K warehouses, $0 < K < |\mathcal{I}|$, such that the resulting supply chain minimizes the sum of investment and expected transportation costs.

We consider a steady-state version of the problem that disregards accumulation or depletion of

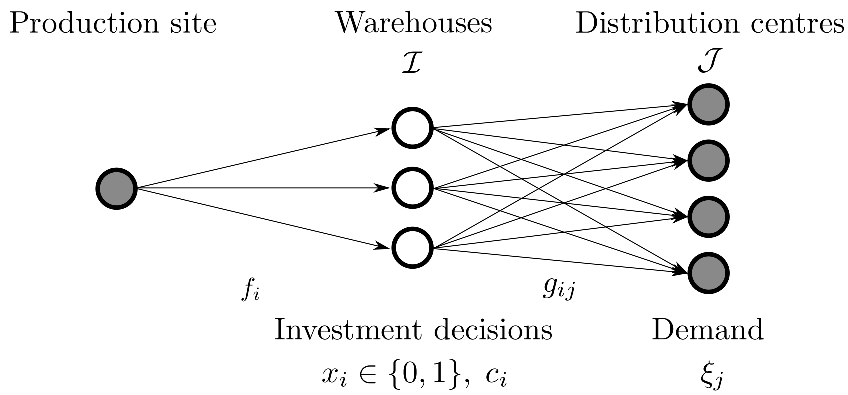

Figure 10. Structure of the supply chain. Shaded nodes represent the existing production site and distribution centers, and unshaded nodes represent candidate locations for warehouses.

stocks. This simplification is motivated in [52, 70, 76], and it allows us to formulate the problem as a two-stage stochastic program. In the first stage of this problem, we determine the location and the capacities of the warehouses. To this end, we define binary variables $x_i \in \{0,1\}$, $i \in \mathcal{I}$, with the interpretation that $x_i = 1$ if a warehouse is built at candidate location i and $x_i = 0$ otherwise. Likewise, we denote the capacity of the warehouse at location $i \in \mathcal{I}$ by c_i. The capacity of each individual warehouse must not exceed $\bar{c}$, and the overall capacity of the warehouses is bounded above by $\overline{C}$. We assume that the construction of warehouse i incurs investment costs $CI_i c_i$ that are proportional to its capacity c_i.

At the beginning of the second stage, we observe the customer demands ξ_j arising at each distribution center $j \in \mathcal{J}$. Since our stochastic program only consists of two stages, we notationally suppress the time stage at which the demands are realized. Once we have observed the customer demands, we select the shipments f_i from the production facility to each warehouse $i \in \mathcal{I}$, as well as the shipments g_{ij} from the warehouses $i \in \mathcal{I}$ to the distribution centers $j \in \mathcal{J}$. We assume that the shipments f_i and g_{ij} incur linear transportation costs $CF_i f_i$ and $CG_{ij} g_{ij}$. Deterministic versions of this supply chain design problem have been studied in [70, 76]. The vast literature on supply chain design is surveyed, amongst others, in [30, 59, 60, 61]. Figure 10 illustrates the structure of the supply chain considered in our problem.

The resulting stochastic program can be formulated as follows.

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i \in \mathcal{I}} CI_i c_i + \mathbb{E}\left[\sum_{i \in \mathcal{I}} CF_i f_i(\xi) + \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} CG_{ij} g_{ij}(\xi) \right] \\
\text{subject to} \quad & \sum_{i \in \mathcal{I}} x_i \leq K, \quad \sum_{i \in \mathcal{I}} c_i \leq \overline{C} \\
& c_i \leq \bar{c}\, x_i && \forall i \in \mathcal{I} \\
& f_i(\xi) \leq c_i, \quad \sum_{j \in \mathcal{J}} g_{ij}(\xi) \leq f_i(\xi) && \forall i \in \mathcal{I} \\
& \sum_{i \in \mathcal{I}} g_{ij}(\xi) \geq \xi_j && \forall j \in \mathcal{J} \\
& x_i \in \{0,1\}, \quad c_i, f_i(\xi), g_{ij}(\xi) \geq 0 && \forall i \in \mathcal{I},\ \forall j \in \mathcal{J}
\end{aligned}
$$

In this problem, we require the constraints to be satisfied for all realizations $\xi \in \Xi$ of the uncertain customer demands. The objective function minimizes the sum of investment and expected transportation costs. The first constraint restricts the total number of warehouses that can be built, as well as the overall capacity of the warehouses. The second constraint ensures that the individual warehouse capacity restrictions are met, and it imposes zero capacities at candidate locations without a warehouse. The third constraint guarantees that the product shipments satisfy flow conservation and the individual warehouse capacities. Finally, the penultimate constraint requires that the customer demands are satisfied at each distribution center. One readily verifies that our supply chain design problem is an instance of the problem $\mathcal{P}_{\text{MI}}$ studied in Section 5.

6.2.1 Numerical Results

We consider an instance of the supply chain design problem where the production facility is in London, while the distribution centers are located in the capitals of 40 European countries. We want to build at most $K = 5$ warehouses in the capitals of the 10 most populous countries, and the warehouse capacities must not exceed the individual and cumulative bounds $\bar{c} = 40$ and $\overline{C} = 100$, respectively. We disregard investment costs ($CI_i = 0$), and we set the transportation costs CF_i and CG_{ij} to 90% and 100% of the geodesic distances between the corresponding cities.

We assume that the expected demand ϱ_j at each distribution center $j \in \mathcal{J}$ is given by the quotient of the corresponding country's population and the overall population of the 40 countries. The actual demand ξ_j at distribution center j is uncertain, and the demand vector ξ follows a uniform distribution with polyhedral support

$$\Xi = \left\{ \xi \in \mathbb{R}_+^{|\mathcal{J}|} : \xi \in [(1-\theta)\cdot 100\varrho,\ (1+\theta)\cdot 100\varrho],\ \ \sum_{j\in\mathcal{J}} \xi_j = 100 \right\},$$

where the parameter $0 \leq \theta \leq 1$ represents the level of uncertainty. Our support definition expresses the view that the cumulative customer demands are known, but the geographical breakdown by country is uncertain. In particular, the demand arising at the capital of country $j \in \mathcal{J}$ can vary within $\theta \cdot 100\%$ of the expected demand ϱ_j.

Figure 11 and Table 3 illustrate the optimal supply chains for $\theta = 0$ (deterministic demand), $\theta = 0.5$ and $\theta = 1$. The figure shows that in the deterministic problem, most of the distribution centers receive their stock from the closest warehouse in order to minimize transportation costs. If the geographic breakdown of the demand is uncertain, however, such an assignment is no longer feasible due to the limited warehouse capacities. Instead, each distribution center potentially receives its stock from different warehouses, and the assignment of warehouses to distribution centers depends on the demand realization $\xi \in \Xi$. This in turn has a profound impact on the design of the optimal supply chain. If the customer demands are deterministic ($\theta = 0$), then the product flows between warehouses and distribution centers are known. In this case, the optimal supply chain design places the warehouses close to the distribution centers in order to take advantage of the cheaper transportation costs CF_i between the production facility and the warehouses. If the geographical breakdown of the customer demands is uncertain, however, such a decentralized warehouse strategy would suffer from costly detours due to the limited warehouse capacities. To illustrate this, consider the warehouse built in Ankara (for $\theta \in \{0, 0.5\}$), which has the second largest expected product demand. If the customer demands are subject to significant geographical uncertainty ($\theta = 1$), then there is a non-negligible chance that a large part of the demand is realized in western European countries. Due to the overall capacity limitation $\overline{C} = 100$, some products would have to be shipped to Turkey before they are delivered to the distribution centers in Western Europe, which would incur high transportation costs. To avoid such detours, the optimal supply chain design adopts a centralized warehouse strategy if customer demands are uncertain.

We solved both instances of the stochastic supply chain design problem ($\theta \in \{0.5, 1\}$) within 10 minutes using linear decision rules. The optimality gap was below 5% ($\theta = 0.5$) and 13% ($\theta = 1$).

Location	$\theta = 0$	$\theta = 0.5$	$\theta = 1$
Berlin	30.3731	0	0
Ankara	11.3988	5.69941	0
Paris	0	19.7618	20
London	40	40	40
Rome	11.3066	0	0
Kiev	0	0	0
Madrid	6.92146	0	0
Warsaw	0	7.58782	0
Bucharest	0	0	0
Amsterdam	0	26.9509	40

Table 3. Installed warehouse capacities for the three uncertainty levels. The cities are ordered according to their expected product demands.

Better results can be obtained by refining the decision rule approximation. The deterministic problem was solved within a few seconds.

References

[1] IBM ILOG CPLEX Homepage. `http://www.ibm.com/software/integration/optimization/cplex-optimizer`.

[2] An, Y., Zeng, B., Zhang, Y., and Zhao, L. Reliable p-median facility location problem: Two-stage robust models and algorithms. *Transportation Research Part B 64* (2014), 54–72.

[3] Ardestani-Jaafari, A., and Delage, E. Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research 64*, 2 (2016), 474–494.

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

[5] Ayoub, J., and Poss, M. Decomposition for adjustable robust linear optimization subject to uncertainty polytope. *Computational Management Science 13*, 2 (2016), 219–239.

[6] Bandi, C., Trichakis, N., and Vayanos, P. Robust multiclass queuing theory for wait time estimation in resource allocation systems. *Management Science* (2017). forthcoming.

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

[8] Ben-Tal, A., Goryashko, A., Guslitzer, E., and Nemirovski, A. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming 99*, 2 (2004), 351–376.

[9] Bertsekas, D. P. *Dynamic Programming and Optimal Control.* Athena Scientific, 2001.

[10] Bertsimas, D., Brown, D. B., and Caramanis, C. Theory and applications of robust optimization. *SIAM Review 53*, 3 (2011), 464–501.

[11] Bertsimas, D., and Caramanis, C. Finite adaptibility in multistage linear optimization. *IEEE Transactions on Automatic Control 55*, 12 (2010), 2751–2766.

[12] Bertsimas, D., and de Ruiter, F. J. C. T. Duality in two-stage adaptive linear optimization: Faster computation and stronger bounds. *INFORMS Journal on Computing 28*, 3 (2016), 500–511.

[13] Bertsimas, D., and Georghiou, A. Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Operations Research 63*, 3 (2015), 610–627.

[14] Bertsimas, D., and Georghiou, A. Binary decision rules for multistage adaptive mixed-integer optimization. *Mathematical Programming 167*, 2 (2018), 395–433.

[15] Bertsimas, D., and Goyal, V. On the power of robust solutions in two-stage stochastic and adaptive optimization problems. *Mathematics of Operations Research 35*, 2 (2010), 284–305.

[16] Bertsimas, D., and Goyal, V. On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical Programming 134*, 2 (2012), 491–531.

[17] Bertsimas, D., Goyal, V., and Sun, X. A. A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization. *Mathematics of Operations Research 36*, 1 (2011), 24–54.

[18] Bertsimas, D., Iancu, D. A., and Parrilo, P. A. Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research 35*, 2 (2010), 363–394.

[19] Bertsimas, D., Iancu, D. A., and Parrilo, P. A. A hierarchy of near-optimal policies for multistage adaptive optimization. *IEEE Transactions on Automatic Control 56*, 12 (2011), 2809–2824.

[20] Bertsimas, D., and Thiele, A. Robust and data-driven optimization: Modern decision making under uncertainty. In *INFORMS TutoRials in Operations Research: Models, Methods, and Applications for Innovative Decision Making*. 2006, pp. 95–122.

[21] Bertsimas, D., and Vayanos, P. Data-driven learning in dynamic pricing using adaptive optimization. Available on Optimization Online, 2017.

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

[23] Campi, M. C., and Garatti, S. A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality. *Journal of Optimization Theory and Applications 148*, 2 (2011), 257–280.

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

[25] Charnes, A., Cooper, W. W., and Symonds, G. H. Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil. *Management Science 4*, 3 (1958), 235–263.

[26] Chen, P., Papageorgiou, L. G., and Pinto, J. M. Medium-term planning of single-stage single-unit multiproduct plants using a hybrid discrete/continuous-time MILP model. *Industrial & Engineering Chemistry Research 47*, 6 (2008), 1925–1934.

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

[28] Chen, X., and Zhang, Y. Uncertain linear programs: Extended affinely adjustable robust counterparts. *Operations Research 57*, 6 (2009), 1469–1482.

[29] Chvátal, V. *Linear Programming.* WH Freeman, 1983.

[30] Daskin, M. S., Snyder, L. V., and Berger, R. T. Facility location in supply chain design. In *Logistics Systems: Design and Optimization*, A. Langevin and D. Riopel, Eds. Springer, 2005, pp. 39–65.

[31] Delage, E., and Iancu, D. A. Robust multistage decision making. *INFORMS TutORials in Operations Research: Models, Methods, and Applications for Innovative Decision Making* (2015), 20–46.

[32] Dupačová, J., Consigli, G., and Wallace, S. Scenarios for multistage stochastic programs. *Annals of Operations Research 100*, 1 (2000), 25–53.

[33] Eisner, M., and Olsen, P. Duality for stochastic programming interpreted as L.P. in L_p-space. *SIAM Journal on Applied Mathematics 28*, 4 (1975), 779–792.

[34] Floudas, C. A., and Lin, X. Mixed integer linear programming in process scheduling: Modeling, algorithms, and applications. *Annals of Operations Research 139*, 1 (2005), 131–162.

[35] Gabrel, V., Murat, C., and Thiele, A. Recent advances in robust optimization: An overview. *European Journal of Operational Research 235*, 3 (2014), 471–483.

[36] Garstka, S. J., and Wets, R. J. B. On decision rules in stochastic programming. *Mathematical Programming 7*, 1 (1974), 117–143.

[37] Georghiou, A., Tsoukalas, A., and Wiesemann, W. Robust dual dynamic programming. Available on Optimization Online, 2016.

[38] Georghiou, A., Tsoukalas, A., and Wiesemann, W. A primal-dual lifting scheme for two-stage robust optimization. Available on Optimization Online, 2017.

[39] Georghiou, A., Wiesemann, W., and Kuhn, D. Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming 152*, 1 (2015), 301–338.

[40] Gilboa, I., and Schmeidler, D. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*, 2 (1989), 141–153.

[41] Goh, J., and Sim, M. Distributionally robust optimization and its tractable approximations. *Operations Research 58*, 4 (2010), 902–917.

[42] Gorissen, B. L., Yanikoğlu, İ., and den Hertog, D. A practical guide to robust optimization. *Omega 53* (2015), 124–137.

[43] Gounaris, C. E., Wiesemann, W., and Floudas, C. A. The robust capacitated vehicle routing problem under demand uncertainty. *Operations Research 61*, 3 (2013), 677–603.

[44] Guslitser, E. Uncertainty-immunized solutions in linear programming. Master's thesis, Technion, 2002.

[45] Hadjiyiannis, M. J., Goulart, P. J., and Kuhn, D. An efficient method to estimate the suboptimality of affine controllers. *IEEE Transactions on Automatic Control 56*, 12 (2011), 2841–2853.

[46] Hadjiyiannis, M. J., Goulart, P. J., and Kuhn, D. A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization. In *Proceedings of*

the 50th IEEE Conference on Decision and Control and European Control Conference (2011), pp. 7386–7391.

[47] Hanasusanto, G. A., and Kuhn, D. Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. *Operations Research* (2017). forthcoming.

[48] Hanasusanto, G. A., Kuhn, D., and Wiesemann, W. K-adaptability in two-stage robust binary programming. *Operations Research 63*, 4 (2015), 877–891.

[49] Hochreiter, R., and Pflug, G. C. Financial scenario generation for stochastic multi-stage decision processes as facility location problems. *Annals of Operations Research 152*, 1 (2007), 257–272.

[50] Kall, P., and Wallace, S. *Stochastic Programming*. Wiley New York, 1994.

[51] Kong, Q., Li, S., Liu, N., Teo, C.-P., and Yan, Z. Robust multi-period vehicle routing under customer order uncertainty. Available on `http://www.columbia.edu/~nl2320/doc/Noshow-MS-1030c.pdf`, 2017.

[52] Kuhn, D., Parpas, P., and Rustem, B. Stochastic optimization of investment planning problems in the electric power industry. In *Process Systems Engineering: Volume 5: Energy Systems Engineering*, M. Georgiadis, E. Kikkinides, and E. Pistikopoulos, Eds. Wiley-VCH, 2008, ch. 4, pp. 215–230.

[53] Kuhn, D., Wiesemann, W., and Georghiou, A. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming 130*, 1 (2011), 177–209.

[54] Lappas, N. H., and Gounaris, C. E. Multi-stage adjustable robust optimization for process scheduling under uncertainty. *AIChE Journal 62*, 5 (2016), 1646–1667.

[55] Liebchen, C., Lübbecke, M., Möhring, R. H., and Stiller, S. The concept of recoverable robustness, linear programming recovery, and railway applications. In *Robust and Online Large-Scale Optimization: Models and Techniques for Transportation Systems*, R. K. Ahuja, R. H. Möhring, and C. D. Zaroliagis, Eds. Springer, 2009, pp. 1–27.

[56] Liu, S., Pinto, J. M., and Papageorgiou, L. G. A TSP-based MILP model for medium-term planning of single-stage continuous multiproduct plants. *Industrial & Engineering Chemistry Research 47*, 20 (2008), 7733–7743.

[57] Lorca, A., and Sun, X. A. Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind. *IEEE Transactions on Power Systems 30*, 4 (2015), 1702–1713.

[58] Lorca, A., and Sun, X. A. Multistage robust unit commitment with dynamic uncertainty sets and energy storage. *IEEE Transactions on Power Systems 32*, 3 (2017), 1678–1688.

[59] Meixell, M. J., and Gargeya, V. B. Global supply chain design: A literature review and critique. *Transportation Research Part E 41*, 6 (2005), 531–550.

[60] Melo, M., Nickel, S., and da Gama, F. S. Facility location and supply chain management – A review. *European Journal of Operational Research 196*, 2 (2009), 401–412.

[61] Min, H., and Zhou, G. Supply chain modeling: Past, present and future. *Computers & Industrial Engineering 43*, 1-2 (2002), 231–249.

[62] Nemirovski, A., and Shapiro, A. Convex approximations of chance constrained programs. *SIAM Journal on Optimization 17*, 4 (2007), 969–996.

[63] Pereira, M. V. F., and Pinto, L. M. V. G. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming 52*, 1 (1991), 359–375.

[64] Pflug, G. C. Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic Constrained Optimization*. Springer, 2000, pp. 272–281.

[65] Pflug, G. C. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming 89*, 2 (2001), 251–271.

[66] Pflug, G. C., and Pichler, A. *Multistage Stochastic Optimization*. Springer, 2014.

[67] Pflug, G. C., and Römisch, W. *Modeling, Measuring and Managing Risk*. World Scientific, 2007.

[68] Rockafellar, R., and Uryasev, S. Optimization of conditional value-at-risk. *Journal of Risk 2* (2000), 21–42.

[69] Ruiz, C., and Conejo, A. J. Robust transmission expansion planning. *European Journal of Operational Research 242*, 2 (2015), 390–401.

[70] Santoso, T., Ahmed, S., Goetschalckx, M., and Shapiro, A. A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research 167*, 1 (2005), 96–115.

[71] Shapiro, A. Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research 209*, 1 (2011), 63–72.

[72] Shapiro, A., Dentcheva, D., and Ruszczyński, A. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2009.

[73] Simchi-Levi, D., Wang, H., and Wei, Y. Constraint generation for two-stage robust network flow problem. Available on Optimization Online, 2018.

[74] Subramanyam, A., Gounaris, C., and Wiesemann, W. K-adaptability in two-stage mixed-integer robust optimization. Available on Optimization Online, 2017.

[75] Subramanyam, A., Mufalli, F., Pinto, J. M., and Gounaris, C. E. Robust multi-period vehicle routing under customer order uncertainty. Available on Optimization Online, 2017.

[76] Tsiakis, P., Shah, N., and Pantelides, C. C. Design of multi-echelon supply chain networks under demand uncertainty. *Industrial & Engineering Chemistry Research 40*, 16 (2001), 3585–3604.

[77] van der Vlerk, M. H. Stochastic programming bibliography. `http://www.eco.rug.nl/mally/spbib.html`.

[78] Vayanos, P., Kuhn, D., and Rustem, B. Decision rules for information discovery in multi-stage stochastic programming. In *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference* (2011), pp. 7368–7373.

[79] Vayanos, P., Kuhn, D., and Rustem, B. A constraint sampling approach for multi-stage robust optimization. *Automatica 48*, 3 (2012), 459–471.

[80] Verderame, P. M., Elia, J. A., Li, J., and Floudas, C. A. Planning and scheduling under uncertainty: A review across multiple sectors. *Industrial & Engineering Chemistry Research 49*, 9 (2010), 3993–4017.

[81] Warrington, J., Goulart, P., Mariéthoz, S., and Morari, M. Policy-based reserves for power systems. *IEEE Transactions on Power Systems 28*, 4 (2013), 4427–4437.

[82] Xu, G., and Burer, S. A copositive approach for two-stage adjustable robust optimization with uncertain right-hand sides. *Computational Optimization and Applications 70*, 1 (2018), 33–59.

[83] Yanikoğlu, İ., Gorissen, B. L., and den Hertog, D. Adjustable robust optimization – a survey and tutorial. Available on ResearchGate, 2017.

[84] Zeng, B., and Zhao, L. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters 41*, 5 (2013), 457–561.

Figure 11. Comparison of the supply chain networks for $\theta = 0$ (top chart), $\theta = 0.5$ (middle chart) and $\theta = 1$ (bottom chart). White lines represent the transportation links between the warehouses and the distribution centers for different realizations of the uncertain demand.