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

Scenario-Free Stochastic Programming with Polynomial Decision Rules

Dimitra Bampou and Daniel Kuhn

Abstract— **Multi-stage stochastic programming provides a versatile framework for optimal decision making under uncertainty, but it gives rise to hard functional optimization problems since the adaptive recourse decisions must be modeled as functions of some or all uncertain parameters. We propose to approximate these recourse decisions by polynomial decision rules and show that the best polynomial decision rule of a fixed degree can be computed efficiently. We also show that the suboptimality of the best polynomial decision rule can be estimated efficiently by solving a dual version of the stochastic program in polynomial decision rules.**

I. INTRODUCTION

Multi-stage stochastic programs have manifold applications in engineering and management science. They naturally arise in power system scheduling, investment planning and supply chain management etc. [1]. Despite their wide applicability, generic multi-stage stochastic programs are computationally intractable, and one has to resort to approximation methods to solve instances of nontrivial sizes. Over the past decades, researchers have mainly devised solution methods that rely on a discretization of the uncertain parameters. Theoretically, these scenario-based approaches can achieve any desired level of accuracy at the cost of proliferating the computational overhead. Recent progress in robust optimization has lead to the emergence of a new class of tractable approximation techniques which preserve the true distribution of the uncertain parameters but restrict the set of recourse decisions to those possessing a specific functional form. Ben-Tal et. al [2] studied *linear* decision rules in the context of robust optimization and proved that the best linear decision rule can be computed efficiently. This tractable upper bound approximation was later extended to the realm of stochastic programming by Shapiro and Nemirovski [3]. To quantify the suboptimality of the best linear decision rule for a given stochastic program, Kuhn et. al [4] proposed to solve its *dual* problem in linear decision rules, which results in an efficiently computable lower bound. A method to improve the approximation quality of linear decision rules was suggested by Chen et al. [5], [6] and Goh and Sim [7], who devised several classes of *piecewise linear* decision rules with desirable scalability properties. The approximation error of piecewise linear decision rules can be estimated by a duality technique due to Georghiou et al. [8]. While piecewise linear decision rules offer a superior approximation quality relative to linear decision rules, they result in an increased computational burden and require tedious fine tuning of multiple design parameters. In the absence of structural information about the true optimal solution of a stochastic program, it may therefore be more appropriate to approximate the recourse decisions by *polynomial* decision rules, which are fully specified by a single design parameter, i.e., their degree. A polynomial decision rule approximation for robust dynamic optimization problems has recently been suggested by Bertsimas et al. [9].

In this paper we assess the potential of polynomial decision rules for solving multistage *stochastic* programs, that is, we impose a polynomial structure on the recourse decision of both the original stochastic program and its dual counterpart. The solutions of the two arising approximate problems provide upper and lower bounds on the true optimal value of the stochastic program, respectively. By using recent results on polynomial optimization and the general problem of moments, we demonstrate that these bounds can be computed in polynomial time by solving two tractable semidefinite programs. A similar approach for simple linear decision rules has been proposed in [4]. First numerical results indicate that even low-degree polynomial decision rules can significantly outperform linear (and even piecewise linear) decision rules.

The rest of the paper is structured as follows. In Sections II and III we develop polynomial decision rule approximations for single and multi-stage stochastic programs, respectively, and in Section IV we assess the potential of our approach for solving a capacity expansion problem from the literature.

Notation Uncertainty is modeled by a probability space $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mathbb{P})$. The elements of the sample space $\mathbb{R}^k$ are denoted by ξ, and the Borel σ-algebra $\mathcal{B}(\mathbb{R}^k)$ represents the set of events that are assigned probabilities by the probability measure $\mathbb{P}$. Let $\mathcal{L}_{k,n}$ denote the space of all Borel measurable functions from $\mathbb{R}^k$ to $\mathbb{R}$ that are bounded on compact sets. Also, denote by $\mathbb{E}(\cdot)$ the expectation operator with respect to $\mathbb{P}$, and let Ξ denote the support of $\mathbb{P}$, i.e., the smallest closed subset of $\mathbb{R}^k$ which has probability 1. The trace of a square matrix is denoted by $\mathrm{tr}(\cdot)$, and $\mathbb{S}^n$ is defined as the space of all symmetric $n \times n$ matrices. For $A, B \in \mathbb{S}^n$, the relation $A \succeq B$ means that $A - B$ is positive semidefinite. We denote by $\mathbb{R}[\xi]$ the ring of real polynomials in ξ and let $\mathbb{R}_d[\xi]$ be its subspace of polynomials of degree at most d. Moreover, we denote by $\mathcal{B}_d(\xi) := (1, \xi_1, \xi_2, \dots, \xi_k, \xi_1^2, \xi_1\xi_2, \dots, \xi_1\xi_k, \xi_2^2, \dots, \xi_k^d)$ the canonical basis of $\mathbb{R}_d[\xi]$, whose dimension is denoted as $s(k,d) := \binom{k+d}{d}$. For $\alpha \in \mathbb{N}_0^k$, ξ^α represents the monomial $\xi_1^{\alpha_1}\xi_2^{\alpha_2}\cdots\xi_k^{\alpha_k}$, and $|\alpha|$ is defined as $\sum_{i=1}^k \alpha_i$. Finally, we define the set $L_d := \left\{\alpha \in \mathbb{N}_0^k : |\alpha| \le d\right\}$. Thus, a polynomial $p \in \mathbb{R}_d[\xi]$ is representable as $p(\xi) = \sum_{\alpha \in L_d} p_\alpha \xi^\alpha =$

This work was supported by EPSRC Grant EP/H0204554/1.

The authors are with the Department of Computing, Imperial College London, {db04,dkuhn}@imperial.ac.uk

$p^\top \mathcal{B}_d(\xi)$. By a slight abuse of notation, we use the same symbol p for a polynomial $p \in \mathbb{R}_d[\xi]$ and the vector of its coefficients $p \in \mathbb{R}^{s(k,d)}$.

II. One-stage Stochastic Programs

We study decision problems under uncertainty in which a decision maker first observes a random vector $\xi \in \mathbb{R}^k$ and then selects a decision $x(\xi) \in \mathbb{R}^n$. The decision $x(\xi)$ must satisfy the inequality constraints $A(\xi)x(\xi) \leq b(\xi)$ and incurs a cost $c(\xi)^\top x(\xi)$. The objective is to choose the function $x \in \mathcal{L}_{k,n}$ so as to minimize the expected cost while satisfying the constraints with probability 1. The functions $x \in \mathcal{L}_{k,n}$ that map observations to decisions are referred to as *decision rules* or *policies*. This decision problem can be formulated as the following one-stage stochastic program.

$$\begin{array}{ll} \inf & \mathbb{E}\big(c(\xi)^\top x(\xi)\big) \\ \text{s.t.} & x \in \mathcal{L}_{k,n} \\ & A(\xi)x(\xi) \leq b(\xi)\ \mathbb{P}\text{-a.s.} \end{array} \qquad (\mathcal{SP})$$

In the remainder we assume that the objective function coefficients and the right hand side function of the constraints depend polynomially on the uncertain parameters ξ, that is, we require that $c(\xi) = C\mathcal{B}_\theta(\xi)$ for some $C \in \mathbb{R}^{n \times s(k,\theta)}$ and $b(\xi) = B\mathcal{B}_\theta(\xi)$ for some $B \in \mathbb{R}^{m \times s(k,\theta)}$. The assumption that both polynomials share the same degree θ is nonrestrictive but simplifies the notation in the rest of the paper. Moreover, we assume that the recourse matrix is independent of ξ, that is, $A(\xi) = A$ for some $A \in \mathbb{R}^{n \times m}$. We emphasize that all results to be developed below could easily be extended to the case of a random recourse matrix with a polynomial dependence on the uncertain parameters. However, this would come at the expense of complicating the notation. The support Ξ of the probability measure $\mathbb{P}$ is a compact basic semi-algebraic set with nonempty interior defined by polynomial inequalities,

$$\Xi = \big\{\xi \in \mathbb{R}^k : w_j(\xi) \geq 0, \quad j = 0, \ldots, J\big\}, \qquad (1)$$

where $w_j \in \mathbb{R}_{d_j}[\xi]$, $j = 0, \ldots, J$. Without loss of generality, we may set $w_0(\xi) = 1$. Furthermore, we assume that there exist polynomials $u_j \in \mathbb{R}[\xi]$ such that the set

$$\{\xi \in \mathbb{R}^n : \textstyle\sum_{j=0}^J u_j(\xi) w_j(\xi) \geq 0\}$$

is compact. This assumption, which is known as Putinar's compactness condition, is nonrestrictive and can always be enforced by appending a dummy constraint $a^2 - \|\xi\|^2 \geq 0$ to the definition of Ξ for some large $a \in \mathbb{R}$, see [10].

The above conditions ensure that problem $\mathcal{SP}$ is well-defined. For the further argumentation it is convenient to introduce a functional slack variable $s \in \mathcal{L}_{k,m}$, which we use to convert the inequality constraints in $\mathcal{SP}$ to equality constraints.

$$\begin{array}{ll} \inf & \mathbb{E}\big(c(\xi)^\top x(\xi)\big) \\ \text{s.t.} & x \in \mathcal{L}_{k,n},\ s \in \mathcal{L}_{k,m} \\ & \left.\begin{array}{l} A(\xi)x(\xi) + s(\xi) = b(\xi) \\ s(\xi) \geq 0 \end{array}\right\}\ \mathbb{P}\text{-a.s.} \end{array} \qquad (2)$$

It is known that problem $\mathcal{SP}$ and its reformulation (2) are #P-hard [11]; there is no efficient algorithm to compute the optimal value of $\mathcal{SP}$ exactly unless P=NP. We thus settle for the modest goal of finding efficiently computable upper and lower bounds. This is achieved by restricting the functional form of the primal and certain dual decision rules in $\mathcal{SP}$ to polynomial functions of a fixed degree. Hence, we reduce the feasible sets of the primal and dual problems. The solutions of these two problems provide upper and lower bounds on the optimal value of $\mathcal{SP}$. We employ recent results on polynomial optimization and the generalized problem of moments to show that these upper and lower bound problems admit tractable approximations.

A. Primal Polynomial Decision Rules

In order to derive a conservative approximation for $\mathcal{SP}$, we reduce the set of admissible decision rules from the space of all measurable functions to the space of polynomial functions of even degree $2d$, that is, we set

$$\begin{array}{lcll} x(\xi) & = & X\mathcal{B}_{2d}(\xi) & \text{for some } X \in \mathbb{R}^{n \times s(k,2d)}, \\ s(\xi) & = & S\mathcal{B}_{2d}(\xi) & \text{for some } S \in \mathbb{R}^{m \times s(k,2d)}. \end{array} \qquad (3)$$

For this restriction to result in a feasible approximate problem we require that $2d \geq \max\{\theta, d_0, \ldots, d_J\}$. Substituting (3) into (2) yields

$$\begin{array}{ll} \inf & \mathrm{tr}\left(T_\theta^\top C^\top X M_{2d}\right) \\ \text{s.t.} & X \in \mathbb{R}^{n \times s(k,2d)},\ S \in \mathbb{R}^{m \times s(k,2d)} \\ & \left.\begin{array}{l} AX\mathcal{B}_{2d}(\xi) + S\mathcal{B}_{2d}(\xi) = BT_\theta\mathcal{B}_{2d}(\xi) \\ S\mathcal{B}_{2d}(\xi) \geq 0 \end{array}\right\}\ \mathbb{P}\text{-a.s.,} \end{array} \qquad (\mathcal{SP}^u)$$

where we use the symmetric moment matrix

$$M_{2d} := \mathbb{E}\big(\mathcal{B}_{2d}(\xi)\mathcal{B}_{2d}(\xi)^\top\big)$$

and a truncation operator $T_\theta : \mathbb{R}^{s(k,2d)} \to \mathbb{R}^{s(k,\theta)}$ which maps the monomial basis $\mathcal{B}_{2d}(\xi)$ to the reduced basis $\mathcal{B}_\theta(\xi)$. As problem $\mathcal{SP}^u$ is obtained by reducing the feasible set of the original stochastic program $\mathcal{SP}$, its optimal value provides indeed an upper bound on the optimal value of $\mathcal{SP}$. Note that $\mathcal{SP}^u$ involves finitely many decision variables, that is, the coefficients X and S of the polynomial decision rules, but infinitely many constraints parametrized by ξ. We first observe that since polynomials are continuous, the almost sure constraints in $\mathcal{SP}^u$ hold in fact for all ξ in the support of $\mathbb{P}$, that is, for all $\xi \in \Xi$. Thus, the equality constraint in $\mathcal{SP}^u$ requires a vector-valued polynomial to vanish identically on a set with nonempty interior. This is possible if and only if all the coefficients of the polynomial vanish. The equality constraint is therefore equivalent to

$$AX + S = BT_\theta.$$

The inequality constraint in $\mathcal{SP}^u$ requires that each component of the vector-valued function $s(\xi)$ belongs to $\mathcal{P}_{2d}(\Xi)$, where $\mathcal{P}_{2d}(\Xi)$ denotes the cone of polynomials of degree $2d$ that are nonnegative on Ξ. For the further argumentation, we define the cone of polynomials of degree $2d$ that have a sums-of-squares (SOS) decomposition relative to Ξ, that is,

$$\Sigma_{2d}(\Xi) := \left\{ s \in \mathbb{R}_{2d}[\xi] \, : \begin{array}{l} s(\xi) = \sum_{j=0}^{J} s_j(\xi) w_j(\xi), \\ s_j \in \Sigma_{2\tilde{d}_j}(\mathbb{R}^k), j = 0, \dots, J \end{array} \right\},$$

where $\tilde{d}_j = d - \lceil \frac{d_j}{2} \rceil$ and

$$\Sigma_{2\tilde{d}_j}(\mathbb{R}^k) := \left\{ s \in \mathbb{R}_{2\tilde{d}_j}[\xi] \, : \begin{array}{l} s(\xi) = \sum_{l=1}^{L} s_l(\xi)^2, \\ \text{for some } s_1, \dots, s_L \in \mathbb{R}[\xi] \end{array} \right\}$$

stands for the cone of SOS polynomials. It is clear that any $s \in \Sigma_{2d}(\Xi)$ is nonnegative on Ξ, and thus $\Sigma_{2d}(\Xi) \subseteq \mathcal{P}_{2d}(\Xi)$. We remark that testing whether a generic polynomial is nonnegative on Ξ (i.e., checking membership in $\mathcal{P}_{2d}(\Xi)$) is NP-hard [12]. However, testing whether a polynomial has an SOS decomposition relative to Ξ (i.e., checking membership in $\Sigma_{2d}(\Xi)$) is equivalent to solving a tractable semidefinite program [13].

Proposition 2.1: Assume that Ξ is defined as in (1). Then, for any $s \in \mathbb{R}_{2d}[\xi]$ the following statements are equivalent.

(i) $s \in \Sigma_{2d}(\Xi)$.

(ii) There exist positive semidefinite matrices $Y^j \in \mathbb{S}^{s(k,\tilde{d}_j)}$, $j = 0, \dots, J$, such that $s = \sum_{j=0}^{J} \Lambda_j^*(Y^j)$, where $\Lambda_j^* : \mathbb{S}^{s(k,\tilde{d}_j)} \to \mathbb{R}^{s(k,2d)}$ is a linear operator defined through

$$\left[\Lambda_j^*(Y^j)\right]_\alpha = \mathrm{tr}\left(Q_\alpha^j Y^j\right), \quad \alpha \in L_{2d}, \tag{4}$$

and $Q_\alpha^j \in \mathbb{S}^{s(k,\tilde{d}_j)}$ is a real symmetric matrix defined through

$$[Q_\alpha^j]_{\beta\gamma} = \begin{cases} [w_j]_\delta & \text{if } \alpha - \beta + \gamma = \delta, \\ 0 & \text{otherwise.} \end{cases}$$

Proof: This result is due to Putinar [10]. The proof is repeated here to keep the paper self-contained. For each $j = 0, \dots, J$ define the linear operators $\Lambda_j : \mathbb{R}^{s(k,2d)} \to \mathbb{S}^{s(k,\tilde{d}_j)}$ through

$$\Lambda_j(\mathcal{B}_{2d}(\xi)) = \sum_{\alpha \in L_{2d}} Q_\alpha^j \xi^\alpha.$$

By construction, we have that

$$\Lambda_j(\mathcal{B}_{2d}(\xi)) = \mathcal{B}_{\tilde{d}_j}(\xi) \mathcal{B}_{\tilde{d}_j}(\xi)^\top w_j^\top \mathcal{B}_{d_j}(\xi)$$

for all $\xi \in \mathbb{R}^k$. The linear operators Λ_j and Λ_j^* are adjoint to each other in the sense that

$$\mathrm{tr}\left(Y^j \Lambda_j(\mathcal{B}_{2d}(\xi))\right) = \Lambda_j^*(Y^j)^\top \mathcal{B}_{2d}(\xi)$$

for all $\xi \in \mathbb{R}^k$ and for all $Y^j \in \mathbb{S}^{s(k,\tilde{d}_j)}$. Assume now that there exist matrices $Y^j \in \mathbb{S}^{s(k,\tilde{d}_j)}$ such that $Y^j \succeq 0$ and $s(\xi) = \sum_{j=0}^{J} \Lambda_j^*(Y^j)^\top \mathcal{B}_{2d}(\xi)$. Thus we have

$$\begin{aligned} s(\xi) &= \sum_{j=0}^{J} \Lambda_j^*(Y^j)^\top \mathcal{B}_{2d}(\xi) \\ &= \sum_{j=0}^{J} \mathrm{tr}\left(Y^j \Lambda_j(\mathcal{B}_{2d}(\xi))\right) \\ &= \sum_{j=0}^{J} \mathrm{tr}\left(Y^j \mathcal{B}_{\tilde{d}_j}(\xi) \mathcal{B}_{\tilde{d}_j}(\xi)^\top w_j^\top \mathcal{B}_{d_j}(\xi)\right) \\ &= \sum_{j=0}^{J} \mathcal{B}_{\tilde{d}_j}(\xi)^\top Y^j \mathcal{B}_{\tilde{d}_j}(\xi) w_j(\xi). \end{aligned}$$

Since all Y^j are positive semidefinite, s is an element of $\Sigma_{2d}(\Xi)$. Thus (ii) implies (i). Conversely, assume that $s \in \Sigma_{2d}(\Xi)$. Thus, $s = \sum_{j=0}^{J} s_j(\xi) w_j(\xi)$ for some polynomials $s_j \in \Sigma_{2\tilde{d}_j}(\mathbb{R}^k)$, $j = 0, \dots, J$. Hence there exist positive semidefinite matrices $Y^j \in \mathbb{S}^{s(k,2\tilde{d}_j)}$ such that $s_j(\xi) = \mathcal{B}_{\tilde{d}_j}(\xi)^\top Y^j \mathcal{B}_{\tilde{d}_j}(\xi)$, $j = 0, \dots, J$, see e.g. [13]. A reversal of the above argument then shows that (i) implies (ii). ■

Proposition 2.1 ensures that $\Sigma_{2d}(\Xi)$ has a manifestly tractable representation as

$$\Sigma_{2d}(\Xi) = \left\{ s \in \mathbb{R}_{2d}[\xi] : \begin{array}{l} s(\xi) = \sum_{j=0}^{J} \Lambda_j^*(Y^j)^\top \mathcal{B}_{2d}(\xi), \\ Y^j \succeq 0, j = 0, \dots, J \end{array} \right\}.$$

For convenience we define $\mathcal{P}_{2d}^m(\Xi)$ and $\Sigma_{2d}^m(\Xi)$ as the sets of all $m \times s(k,2d)$-matrices whose rows are all elements of $\mathcal{P}_{2d}(\Xi)$ and $\Sigma_{2d}(\Xi)$, respectively. The inequality constraint in $\mathcal{SP}^u$ is equivalent to $S \in \mathcal{P}_{2d}^m(\Xi)$, and therefore $\mathcal{SP}^u$ is generally intractable. To overcome this deficiency, we approximate the inequality constraint by $S \in \Sigma_{2d}^m(\Xi)$, which yields the following approximate problem.

$$\begin{array}{ll} \inf & \mathrm{tr}\left(T_\theta^\top C^\top X M_{2d}\right) \\ \text{s.t.} & X \in \mathbb{R}^{n \times s(k,2d)}, \quad S \in \mathbb{R}^{m \times s(k,2d)} \\ & AX + S = BT_\theta \\ & S \in \Sigma_{2d}^m(\Xi) \end{array} \tag{$\tilde{\mathcal{SP}}^u$}$$

By construction, $\tilde{\mathcal{SP}}^u$ represents a conservative approximation for $\mathcal{SP}^u$. Our insights can be summarized as follows.

Theorem 2.2: We have $\inf \tilde{\mathcal{SP}}^u \geq \inf \mathcal{SP}^u \geq \inf \mathcal{SP}$, and the approximate problem $\tilde{\mathcal{SP}}^u$ is computationally tractable, that is, it can be solved in polynomial time.

B. Dual Polynomial Decision Rules

Similar techniques to those used in Section II-A can be applied to the dual of the stochastic program $\mathcal{SP}$. This will allow us to construct a computationally tractable lower bound on the optimal value of $\mathcal{SP}$. A related approach has been proposed in [4] to estimate the suboptimality of *linear* decision rules.

For ease of exposition, denote by '$\inf_{x,s}$' the infimum operator over all $x \in \mathcal{L}_{k,n}$ and over all $s \in \mathcal{L}_{k,m}$ that are almost surely nonnegative, by '$\sup_y$' the supremum operator over all $y \in \mathcal{L}_{k,m}$ and by '$\sup_Y$' the supremum operator over all $Y \in \mathbb{R}^{m \times s(k,2d)}$. We first introduce a min-max reformulation of problem (2) in which the equality constraints are dualized.

$$\begin{array}{ll} \inf\limits_{x,s} \sup\limits_{y} & \mathbb{E}\big(c(\xi)^\top x(\xi) + y(\xi)^\top \left[Ax(\xi) + s(\xi) - b(\xi)\right]\big) \\ \text{s.t.} & s(\xi) \geq 0 \ \mathbb{P}\text{-a.s.} \end{array} \tag{5}$$

Problems (2) and (5) are indeed equivalent since the maximization over the dual decision rules $y \in \mathcal{L}_{k,m}$ imposes an infinite penalty on every primal solution $(x, s) \in \mathcal{L}_{k,n} \times \mathcal{L}_{k,m}$ which violates the equality constraints $Ax(\xi) + s(\xi) = b(\xi)$ on a set of strictly positive probability. Using the equivalence of (2) and (5) we obtain

$$
\begin{aligned}
&\inf_{x,s} \; \mathcal{SP} \\
&= \inf_{x,s} \; \sup_{y} \; \mathbb{E}\big(c(\xi)^\top x(\xi) + y(\xi)^\top [Ax(\xi) + s(\xi) - b(\xi)]\,\big) \\
&\geq \inf_{x,s} \; \sup_{Y} \; \mathbb{E}\big(c(\xi)^\top x(\xi) \cdots \\
&\qquad\qquad \cdots + (Y\mathcal{B}_{2d}(\xi))^\top \left[Ax(\xi) + s(\xi) - b(\xi)\right]\big).
\end{aligned}
$$

In the last expression we require the dual decision rules to be representable as $y(\xi) = Y\mathcal{B}_{2d}(\xi)$ for some matrix $Y \in \mathbb{R}^{m \times s(k,2d)}$. Thus we restrict the dual feasible set to contain only polynomial decision rules of even degree $2d$, where we require again that $2d \geq \max\{\theta, d_0, \ldots, d_J\}$. The inner maximization in the third line can be carried out explicitly to yield

$$
\begin{array}{ll}
\inf & \mathbb{E}\big(c(\xi)^\top x(\xi)\big) \\
\text{s.t.} & x \in \mathcal{L}_{k,n},\; s \in \mathcal{L}_{k,m} \\
& \left.\begin{array}{l} \mathbb{E}\big(\,[Ax(\xi) + s(\xi) - b(\xi)]\,\mathcal{B}_{2d}(\xi)^\top\big) = 0 \\ s(\xi) \geq 0 \end{array}\right\} \mathbb{P}\text{-a.s.}
\end{array}
\tag{$\mathcal{SP}^l$}
$$

Notice that any (x, s) feasible in (2) satisfies $Ax(\xi)+s(\xi)-b(\xi) = 0$ for $\mathbb{P}$-almost all ξ and thus will satisfy the less restrictive expectation constraint in $\mathcal{SP}^l$. This observation confirms that problem $\mathcal{SP}^l$ is a relaxation of (2) and that its optimal value provides a lower bound on the optimal value of $\mathcal{SP}$. Problem $\mathcal{SP}^l$ involves only finitely many equality constraints, but it involves a continuum of decision variables and inequality constraints. In the remainder of this section we will show that $\mathcal{SP}^l$ admits a tractable lower bound approximation. Our reasoning relies on the following technical results about the symmetric moment matrix M_{2d}.

Proposition 2.3: M_{2d} is positive definite and invertible.

Proof: By definition, M_{2d} is positive semidefinite. Assume now that there exists a $v \in \mathbb{R}^{s(k,2d)}$ such that

$$
v^\top M_{2d} v = 0 \iff \mathbb{E}\left(\left(v^\top \mathcal{B}_{2d}(\xi)\right)^2\right) = 0.
$$

This implies that the polynomial $v^\top \mathcal{B}_{2d}(\xi)$ vanishes identically on Ξ. Since Ξ has nonempty interior, we conclude that $v = 0$. Hence M_{2d} is positive definite and, a fortiori, invertible. ■

Next, we define new decision variables $X \in \mathbb{R}^{n\times s(k,2d)}$ and $S \in \mathbb{R}^{m\times s(k,2d)}$ in problem $\mathcal{SP}^l$, which are uniquely determined by the decision rules $x \in \mathcal{L}_{k,n}$ and $s \in \mathcal{L}_{k,m}$, respectively, through the new constraints

$$
\begin{array}{rcl}
XM_{2d} & = & \mathbb{E}\big(x(\xi)\mathcal{B}_{2d}(\xi)^\top\big), \\
SM_{2d} & = & \mathbb{E}\big(s(\xi)\mathcal{B}_{2d}(\xi)^\top\big).
\end{array}
\tag{6}
$$

We can use the relations (6) to re-express the objective function of $\mathcal{SP}^l$ as $\mathrm{tr}(T_\theta^\top C^\top X M_{2d})$, where the truncation operator T_θ is defined as in Section II-A. Moreover, substituting (6) into the expectation constraints of $\mathcal{SP}^l$ yields $AXM_{2d} + SM_{2d} = BT_\theta M_{2d}$, which is equivalent to $AX + S = BT_\theta$ since M_{2d} is invertible. Thus we can reformulate $\mathcal{SP}^l$ as

$$
\begin{array}{ll}
\inf & \mathrm{tr}\left(T_\theta^\top C^\top X M_{2d}\right) \\
\text{s.t.} & X \in \mathbb{R}^{n\times s(k,2d)},\; S \in \mathbb{R}^{m\times s(k,2d)} \\
& AX + S = BT_\theta \\
& \exists x \in \mathcal{L}_{k,n} : XM_{2d} = \mathbb{E}\big(x(\xi)\mathcal{B}_{2d}(\xi)^\top\big) \\
& \exists s \in \mathcal{L}_{k,m} : SM_{2d} = \mathbb{E}\big(s(\xi)\mathcal{B}_{2d}(\xi)^\top\big), \\
& \qquad\qquad s(\xi) \geq 0 \;\mathbb{P}\text{-a.s.}
\end{array}
\tag{7}
$$

The penultimate constraint in (7) is redundant and can be omitted without affecting the problem's feasible set. Indeed, for any $X \in \mathbb{R}^{n\times (k,2d)}$ the polynomial decision rule $x(\xi) = X\mathcal{B}_{2d}(\xi) \in \mathcal{L}_{k,n}$ satisfies the postulated conditions. However, the last constraint involves the solution of m multidimensional moment problems.

For the further argumentation, we introduce the cone $\mathcal{M}_{2d}(\Xi)$ of moment sequences with a representing measure supported on Ξ, that is, we set

$$
\mathcal{M}_{2d}(\Xi) := \left\{ y \in \mathbb{R}^{s(k,2d)} : \begin{array}{l} \exists \mu \in \mathcal{N} \text{ with} \\ y = \int_\Xi \mathcal{B}_{2d}(\xi)\mu(\mathrm{d}\xi) \end{array} \right\},
$$

where $\mathcal{N}$ denotes the set of nonnegative Borel measures supported on Ξ. Moreover, we introduce the cone

$$
\mathcal{M}^+_{2d}(\Xi) = \left\{ y \in \mathbb{R}^{s(k,2d)} \;:\; \Lambda_j(y) \succeq 0 \quad j = 0, \ldots, J \right\},
$$

where the matrix-valued functions Λ_j are defined as in the proof of Proposition 2.1.

Proposition 2.4: The cones $\mathcal{M}_{2d}(\Xi)$ and $\mathcal{M}^+_{2d}(\Xi)$ satisfy the following relations.

(i) $\mathcal{P}_{2d}(\Xi)$ and $\mathcal{M}_{2d}(\Xi)$ are dual to each other.
(ii) $\Sigma_{2d}(\Xi)$ and $\mathcal{M}^+_{2d}(\Xi)$ are dual to each other.
(iii) $\mathcal{M}_{2d}(\Xi) \subseteq \mathcal{M}^+_{2d}(\Xi)$.

Proof: This result is due to Haviland [14]. The proof is repeated here to keep the paper self-contained. (i) The cones $\mathcal{P}_{2d}(\Xi)$ and $\mathcal{M}_{2d}(\Xi)$ are dual to each other if and only if $\mathcal{P}_{2d}(\Xi) = (\mathcal{M}_{2d}(\Xi))^*$ and $(\mathcal{P}_{2d}(\Xi))^* = \mathrm{cl}\,\mathcal{M}_{2d}(\Xi)$. For $p \in \mathcal{P}_{2d}(\Xi)$ and $y \in \mathcal{M}_{2d}(\Xi)$ we have

$$
p^\top y = \int_\Xi p^\top \mathcal{B}_{2d}(\xi)\mu(\mathrm{d}\xi) = \int_\Xi p(\xi)\mu(\mathrm{d}\xi) \geq 0
$$

for some $\mu \in \mathcal{N}$. By the definition of cone duality we may thus conclude that $\mathcal{P}_{2d}(\Xi) \subseteq (\mathcal{M}_{2d}(\Xi))^*$. Suppose now that $p \notin \mathcal{P}_{2d}(\Xi)$, which implies that there exists a $\xi_0 \in \Xi$ such that $p(\xi_0) < 0$. Let μ be the Dirac measure concentrated on the set $\{\xi_0\}$ and let $y = \mathcal{B}_{2d}(\xi_0)$ be its sequence of moments. Then, we find

$$
p^\top y = \int_\Xi p^\top \mathcal{B}_{2d}(\xi)\mu(\mathrm{d}\xi) = p(\xi_0) < 0.
$$

Thus, p is not contained in $(\mathcal{M}_{2d}(\Xi))^*$, which implies that $(\mathcal{M}_{2d}(\Xi))^* \subseteq \mathcal{P}_{2d}(\Xi)$. Hence, $\mathcal{P}_{2d}(\Xi) = (\mathcal{M}_{2d}(\Xi))^*$. As $\mathcal{M}_{2d}(\Xi)$ is convex and closed, this identity further implies that $(\mathcal{P}_{2d}(\Xi))^* = (\mathcal{M}_{2d}(\Xi))^{**} = \mathcal{M}_{2d}(\Xi)$.

(ii) By cone duality, we have

$$\begin{aligned}
& y \in (\Sigma_{2d}(\Xi))^* \\
\iff & p^\top y \geq 0 && \forall p \in \Sigma_{2d}(\Xi) \\
\iff & \sum_{j=0}^{J} \Lambda_j^*(Y^j)^\top y \geq 0 && \left.\begin{array}{l} \forall Y^j \succeq 0, \\ \quad j = 0, \dots, J \end{array}\right. \\
\iff & \operatorname{tr}(Y^j \Lambda_j(y)) \geq 0 && \\
\iff & \Lambda_j(y) \succeq 0 && \forall j = 0, \dots, J.
\end{aligned}$$

Hence, $\mathcal{M}_{2d}^+(\Xi)$ is the dual cone of $\Sigma_{2d}(\Xi)$. As $\Sigma_{2d}(\Xi)$ is convex and closed, this further implies that $(\mathcal{M}_{2d}(\Xi))^* = (\Sigma_{2d}(\Xi))^{**} = \Sigma_{2d}(\Xi)$.

(iii) The known inclusion $\mathcal{P}_{2d}(\Xi) \supseteq \Sigma_{2d}(\Xi)$ implies via assertions (i) and (ii) that $\mathcal{M}_{2d}(\Xi) = (\mathcal{P}_{2d}(\Xi))^* \subseteq (\Sigma_{2d}(\Xi))^* = \mathcal{M}_{2d}^+(\Xi)$. ■

Inspecting problem (7), we see that the last constraint requires each component $s_i(\xi)$, $i = 1, \dots, m$, of the vector-valued function $s(\xi)$ to be the density function of a measure $\mu_i \in \mathcal{N}$ whose moments coincide with the ith row of SM_{2d}. This implies via Proposition 2.4 that the ith row of SM_{2d} must be contained in $\mathcal{M}_{2d}(\Xi) \subseteq \mathcal{M}_{2d}^+(\Xi)$. For the further argumentation, define $\mathcal{M}_{2d}^m(\Xi)$ and $\mathcal{M}_{2d}^{m+}(\Xi)$ as the cones of all $m \times s(k, 2d)$-matrices whose rows are all contained in $\mathcal{M}_{2d}(\Xi)$ and $\mathcal{M}_{2d}^+(\Xi)$, respectively. The above reasoning implies that we obtain a tractable relaxation for problem $\mathcal{SP}^l$ if we replace the last existence constraint in (7) by the requirement $SM_{2d} \in \mathcal{M}_{2d}^{m+}(\Xi)$.

$$\begin{array}{ll}
\inf & \operatorname{tr}\left(T_\theta^\top C^\top X M_{2d}\right) \\
\text{s.t.} & X \in \mathbb{R}^{n \times s(k,2d)},\ S \in \mathbb{R}^{m \times s(k,2d)} \\
& AX + S = BT_\theta \\
& SM_{2d} \in \mathcal{M}_{2d}^{m+}(\Xi)
\end{array} \qquad (\tilde{\mathcal{SP}}^l)$$

By construction, $\tilde{\mathcal{SP}}^l$ represents a relaxation of $\mathcal{SP}^l$. This result culminates in the following theorem.

Theorem 2.5: We have $\inf \tilde{\mathcal{SP}}^l \leq \inf \mathcal{SP}^l \leq \inf \mathcal{SP}$, and the approximate problem $\tilde{\mathcal{SP}}^l$ is computationally tractable.

III. Multi-stage Stochastic Programs

We now demonstrate how the polynomial decision rule approximations developed for one-stage stochastic programs can be extended to multi-stage problems of the form

$$\begin{array}{ll}
\inf & \mathbb{E}\left(\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)\right) \\
\text{s.t.} & x_t \in \mathcal{L}_{k^t, n_t}\ \forall t \in \mathbb{T} \\
& \sum_{s=1}^{t} A_{ts}(\xi^t) x_s(\xi^s) \leq b_t(\xi^t)\ \mathbb{P}\text{-a.s.}\ \forall t \in \mathbb{T}.
\end{array} \qquad (\mathcal{MSP})$$

We continue to work with the probability space $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mathbb{P})$ and assume that $\mathbb{P}$ is supported on a compact basic semi-algebraic set Ξ of the type (1) with nonempty interior. Moreover, we assume that the elements of the sample space are now representable as $\xi = (\xi_1, \dots, \xi_T)$, where the subvectors $\xi_t \in \mathbb{R}^{k_t}$ are observed sequentially at time points indexed by $t \in \mathbb{T} := \{1, \dots, T\}$. The history of observations up to time t is denoted by $\xi^t := (\xi_1, \dots, \xi_t) \in \mathbb{R}^{k^t}$, where $k^t := \sum_{s=1}^{t} k_s$. For consistency, we require that $\xi^T = \xi$ and $k^T = k$. We use $\mathbb{E}_t(\cdot)$ to denote conditional expectation with respect to $\mathbb{P}$ given the random variable ξ^t. Finally, we introduce truncation operators $P_{d,t} : \mathbb{R}^{s(k,d)} \to \mathbb{R}^{s(k^t,d)}$ for any $t \in \mathbb{T}$ and $d \in \mathbb{N}_0$ that map the monomial basis $\mathcal{B}_d(\xi)$ to the reduced basis $\mathcal{B}_d(\xi^t)$.

The decision $x_t(\xi^t)$ is selected at time t after the outcome history ξ^t has been observed but before the future outcomes $\{\xi_s\}_{s>t}$ have been revealed. The objective is to find an optimal sequence of decision rules $x_t \in \mathcal{L}_{k^t, n_t}$, $t \in \mathbb{T}$, which map the available observations to decisions and minimize a linear expected cost function subject to linear constraints. The requirement that x_t depends only on ξ^t reflects the causality of the decision process.

Without much loss of generality, we assume henceforth that $\mathcal{MSP}$ has fixed recourse and that the objective function coefficients and the right-hand side vectors are non-anticipative polynomial functions of the uncertain parameters. For notational convenience, we assume that all these polynomials share the same degree θ. Thus, we postulate that $c_t(\xi^t) = C_t P_{\theta,t} \mathcal{B}_\theta(\xi)$ for some $C_t \in \mathbb{R}^{n_t \times s(k^t,\theta)}$, $b_t(\xi^t) = B_t P_{\theta,t} \mathcal{B}_\theta(\xi)$ for some $B_t \in \mathbb{R}^{m_t \times s(k^t,\theta)}$ and that the matrices $A_{ts}(\xi^t) = A_{ts} \in \mathbb{R}^{m_t \times n_s}$ are independent of ξ.

By introducing a sequence of non-anticipative slack variables $s_t \in \mathcal{L}_{k^t, m_t}$, $t \in \mathbb{T}$, we can reduce $\mathcal{MSP}$ to the following standard form.

$$\begin{array}{ll}
\inf & \mathbb{E}\left(\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)\right) \\
\text{s.t.} & x_t \in \mathcal{L}_{k^t, n_t},\ s_t \in \mathcal{L}_{k^t, m_t}\ \forall t \in \mathbb{T} \\
& \left.\begin{array}{rcl} \sum_{s=1}^{t} A_{ts} x_s(\xi^s) + s_t(\xi^t) & = & b_t(\xi^t) \\ s_t(\xi^t) & \geq & 0 \end{array}\right\} \begin{array}{l} \mathbb{P}\text{-a.s.} \\ \forall t \in \mathbb{T}. \end{array}
\end{array} \qquad (8)$$

Problem $\mathcal{MSP}$ is generically computationally intractable [3]. However, as in the single-stage case, it can be approximated from above and below by two semi-infinite problems $\mathcal{MSP}^u$ and $\mathcal{MSP}^l$, respectively, which are obtained by restricting the primal and certain dual decision rules to those that are representable as polynomial functions of the uncertain parameters ξ. For a fixed degree of the polynomial approximations, these semi-infinite problems can be approximated by tractable semidefinite programs.

Problem $\mathcal{MSP}^u$ is obtained by solving the original problem $\mathcal{MSP}$ in polynomial decision rules of degree $2d \geq \max\{\theta, d_0, \dots, d_J\}$. The decision and slack variables can thus be written as $x_t(\xi^t) = X_t P_{2d,t} \mathcal{B}_{2d}(\xi)$ for some $X_t \in \mathbb{R}^{n_t \times s(k^t,2d)}$ and $s_t(\xi^t) = S_t P_{2d,t} \mathcal{B}_{2d}(\xi)$ for some $S_t \in \mathbb{R}^{m_t \times s(k^t,2d)}$. To ensure that this approximation leads to a tractable problem, we require that $\mathbb{E}_t(\mathcal{B}_{2d}(\xi))$ is essentially polynomial in ξ^t, that is, $\mathbb{E}_t(\mathcal{B}_{2d}(\xi)) = M_t P_{2d,t} \mathcal{B}_{2d}(\xi)$ $\mathbb{P}$-a.s. for some matrix $M_t \in \mathbb{R}^{s(k,2d) \times s(k^t,2d)}$ for all $t \in \mathbb{T}$. Using the truncation operator T_θ and the second order moment matrix M_{2d} defined in Section II and the fact that Ξ has nonempty interior, problem $\mathcal{MSP}^u$ can be approximated by the following semidefinite program.

$$\begin{array}{ll} \inf & \sum_{t=1}^{T} \operatorname{tr}\left(C_t P_{\theta,t} T_\theta M_{2d} P_{2d,t}^\top X_t^\top\right) \\ \text{s.t.} & X_t \in \mathbb{R}^{n_t \times s(k^t,2d)},\ S_t \in \mathbb{R}^{m_t \times s(k^t,2d)} \\ & \left.\begin{array}{l} \sum_{s=1}^{t} A_{ts} X_s P_{2d,s} + S_t P_{2d,t} = B_t P_{\theta,t} T_\theta \\ S_t P_{2d,t} \in \Sigma_{2d}^m(\Xi) \end{array}\right\} \forall t \in \mathbb{T} \end{array} \qquad (\mathcal{M}\tilde{\mathcal{S}}\mathcal{P}^u)$$

By construction, $\mathcal{M}\tilde{\mathcal{S}}\mathcal{P}^u$ constitutes a tractable conservative approximation for $\mathcal{MSP}^u$.

Theorem 3.1: We have $\inf\ \mathcal{M}\tilde{\mathcal{S}}\mathcal{P}^u \geq \inf\ \mathcal{MSP}^u \geq \inf\ \mathcal{MSP}$, and the approximate problem $\mathcal{M}\tilde{\mathcal{S}}\mathcal{P}^u$ is computationally tractable.

Next, we aim at estimating the degree of suboptimality of the best polynomial policy obtained from problem $\mathcal{M}\tilde{\mathcal{S}}\mathcal{P}^u$. To this end, we reexpress the standardized stochastic program (8) as a min-max problem in which the dual variable of the t-th equality constraint is given by a non-anticipative decision rule $y_t \in \mathcal{L}_{k^t,m_t}$, $t \in \mathbb{T}$. To obtain a lower bound on the optimal value of $\mathcal{MSP}$ we require these dual decision rules to be representable as polynomials of degree $2d$, i.e., we set $y_t(\xi^t) = Y_t P_{2d,t} \mathcal{B}_{2d}(\xi)$ for some $Y_t \in \mathbb{R}^{m_t \times s_t(k^t,2d)}$ for all $t \in \mathbb{T}$. Carrying out the inner maximization over the variables $\{Y_t\}_{t\in\mathbb{T}}$ we obtain the following semi-infinite program.

$$\begin{array}{ll} \inf & \mathbb{E}\left(\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)\right) \\ \text{s.t.} & x_t \in \mathcal{L}_{k^t,n_t},\ s_t \in \mathcal{L}_{k^t,m_t}\ \forall t \in \mathbb{T} \\ & \mathbb{E}\Big(\Big[\sum_{s=1}^{t} A_{ts} x_s(\xi^s) + s_t(\xi^t) - b_t(\xi^t)\Big] \cdots \\ & \qquad \cdots \times \mathcal{B}_{2d}(\xi)^\top P_{2d,t}^\top\Big) = 0\ \forall t \in \mathbb{T} \\ & s_t(\xi^t) \geq 0 \quad \mathbb{P}\text{-a.s. } \forall t \in \mathbb{T} \end{array} \qquad (\mathcal{MSP}^l)$$

It is easy to verify that any (x, s) feasible in (8) is also feasible in $\mathcal{MSP}^l$ with the same objective value. Hence, $\mathcal{MSP}^l$ is indeed a relaxation of $\mathcal{MSP}$. To obtain a tractable approximation for $\mathcal{MSP}^l$, we employ similar arguments as in Section II-B. In particular, we introduce new decision variables $X_t \in \mathbb{R}^{n_t \times s(k^t,2d)}$ and $S_t \in \mathbb{R}^{m_t \times s(k^t,2d)}$ which are uniquely determined by the decision rules $x_t \in \mathcal{L}_{k^t,n_t}$ and $s_t \in \mathcal{L}_{k^t,m_t}$, respectively, through the new constraints

$$\begin{array}{rcl} X_t P_{2d,t} M_{2d} &=& \mathbb{E}\big(x_t(\xi^t)\mathcal{B}_{2d}(\xi)^\top\big), \\ S_t P_{2d,t} M_{2d} &=& \mathbb{E}\big(s_t(\xi^t)\mathcal{B}_{2d}(\xi)^\top\big). \end{array} \tag{9}$$

In the following lemma we demonstrate that the new constraints (9) do not restrict the choice of x_t and s_t.

Lemma 3.2: For any given $x_t \in \mathcal{L}_{k^t,n_t}$ and $s_t \in \mathcal{L}_{k^t,m_t}$ there exist unique matrices X_t and S_t satisfying (9).

Proof: Define $X_t \in \mathbb{R}^{n_t \times s(k^t,2d)}$ through

$$\mathbb{E}(x_t(\xi^t)\mathcal{B}_{2d}(\xi)^\top) P_{2d,t}^\top = X_t P_{2d,t} M_{2d} P_{2d,t}^\top. \tag{10}$$

Notice that $P_{2d,t} M_{2d} P_{2d,t}^\top$ is a principal submatrix of M_{2d}, and hence it is invertible. This implies that X_t is uniquely determined by (10). Recall that $\mathbb{E}_t(\mathcal{B}_{2d}(\xi)) = M_t P_{2d,t} \mathcal{B}_{2d}(\xi)$ $\mathbb{P}$-a.s. Thus, we conclude that

$$\begin{array}{rl} & \mathbb{E}\left(x_t(\xi^t)\mathcal{B}_{2d}(\xi)^\top\right) \\ = & \mathbb{E}\big(x_t(\xi^t)\mathbb{E}_t(\mathcal{B}_{2d}(\xi))^\top\big) \\ = & \mathbb{E}\left(x_t(\xi^t)\mathcal{B}_{2d}(\xi)^\top\right) P_{2d,t}^\top M_t^\top \\ = & X_t P_{2d,t} M_{2d} P_{2d,t}^\top M_t^\top \\ = & X_t \mathbb{E}\left(P_{2d,t}\mathcal{B}_{2d}(\xi)\mathcal{B}_{2d}(\xi)^\top P_{2d,t}^\top\right) M_t^\top \\ = & X_t \mathbb{E}\left(P_{2d,t}\mathcal{B}_{2d}(\xi)\mathbb{E}_t\left(\mathcal{B}_{2d}(\xi)^\top\right)\right) \\ = & X_t P_{2d,t} \mathbb{E}\left(\mathcal{B}_{2d}(\xi)\mathcal{B}_{2d}(\xi)^\top\right) \\ = & X_t P_{2d,t} M_{2d}. \end{array}$$

Hence, X_t defined through (10) satisfies (9). A matrix S_t satisfying the second relation in (9) can be constructed in a similar manner. ■

If we replace the decision rules x_t and s_t in problem $\mathcal{MSP}^l$ with the finite dimensional variables X_t and S_t, the following existence constraints appear.

$$\exists x_t \in \mathcal{L}_{k^t,n_t} : X_t P_{2d,t} M_{2d} = \mathbb{E}\left(x_t(\xi^t)\mathcal{B}_{2d}(\xi)^\top\right) \tag{11a}$$

$$\begin{array}{l} \exists s_t \in \mathcal{L}_{k^t,m_t} : S_t P_{2d,t} M_{2d} = \mathbb{E}\left(s_t(\xi^t)\mathcal{B}_{2d}(\xi)^\top\right), \\ \qquad s_t(\xi^t) \geq 0\ \mathbb{P}\text{-a.s.} \end{array} \tag{11b}$$

Constraint (11a) is redundant and can be omitted without affecting the problem's feasible set. Indeed, for any matrix $X_t \in \mathbb{R}^{n_t \times s(k^t,s2d)}$, the polynomial decision rule $x_t(\xi^t) = X_t P_{2d,t} \mathcal{B}_{2d}(\xi) \in \mathcal{L}_{k^t,n_t}$ satisfies the postulated condition (11a). To obtain a tractable relaxation to constraint (11b) we use Proposition 2.4, which is applicable due to the following lemma.

Lemma 3.3: For any given $S_t \in \mathbb{R}^{m_t \times s(k^t,2d)}$ constraint (11b) is equivalent to

$$\begin{array}{l} \exists \tilde{s}_t \in \mathcal{L}_{k,m_t} : S_t P_{2d,t} M_{2d} = \mathbb{E}\left(\tilde{s}_t(\xi)\mathcal{B}_{2d}(\xi)^\top\right), \\ \qquad \tilde{s}_t(\xi) \geq 0\ \mathbb{P}\text{-a.s.} \end{array} \tag{11c}$$

Proof: It is clear that (11b) implies the less restrictive condition (11c). Assume now that (11c) holds and define $s_t(\xi^t) = \mathbb{E}_t(\tilde{s}_t(\xi))$. Then, we find

$$\begin{array}{rcl} \mathbb{E}\left(s_t(\xi^t)\mathcal{B}_{2d}(\xi)^\top\right) &=& \mathbb{E}\left(\mathbb{E}_t\left(\tilde{s}_t(\xi)\right)\mathbb{B}_{2d}(\xi)\right) \\ &=& \mathbb{E}\left(\tilde{s}_t(\xi)\mathcal{B}_{2d}(\xi)^\top\right) P_{2d,t}^\top M_t^\top \\ &=& S_t P_{2d,t} M_{2d}, \end{array}$$

where the last equality follows from Lemma 3.2. ■

Using the above results and the fact that the moment matrix M_{2d} is invertible, $\mathcal{MSP}^l$ can be approximated by the following tractable semidefinite program

$$\begin{array}{ll} \inf & \sum_{t=1}^{T} \operatorname{tr}\left(C_t P_{\theta,t} T_\theta M_{2d} P_{2d,t}^\top X_t^\top\right) \\ \text{s.t.} & X_t \in \mathbb{R}^{n_t \times s(k^t,2d)},\ S_t \in \mathbb{R}^{m_t \times s(k^t,2d)}\ \forall t \in \mathbb{T} \\ & \sum_{s=1}^{t} A_{ts} X_s P_{2d,s} P_{2d,t}^\top + S_t P_{2d,t} P_{2d,t}^\top \cdots \\ & \qquad \cdots = B_t P_{\theta,t} T_\theta P_{2d,t}^\top\ \forall t \in \mathbb{T} \\ & S_t P_{2d,t} M_{2d} \in \mathcal{M}_{2d}^{m+}(\Xi)\ \forall t \in \mathbb{T}. \end{array} \qquad (\mathcal{M}\tilde{\mathcal{S}}\mathcal{P}^l)$$

By construction, problem $\mathcal{M}\tilde{\mathcal{S}}\mathcal{P}^l$ represents a tractable progressive approximation for $\mathcal{MSP}^l$.

TABLE I

INPUT PARAMETERS

Parameter	Value	Random Variable	Range
$\bar{g}_n$	$3.5, \forall n \in N$	δ_1	[0.3,1.5]
$\bar{f}_m$	$3.5, \forall m \in M$	δ_2	[0.36,1.8]
c_1	1	δ_3	[0.42,2.1]
c_2	0.4	δ_4	[0.48,2.4]
c_3	1.5	δ_5	[0.54,2.7]
d_1	5	ζ_1	[0.2,1]
d_2	0.2	ζ_2	[0.2,0.5]
d_3	0.4	ζ_3	[1,2]
d_4	0.6		
d_5	0.1		

Theorem 3.4: We have $\inf\ \mathcal{M}\tilde{\mathcal{SP}}^l \leq \inf\ \mathcal{MSP}^l \leq \inf\ \mathcal{MSP}$, and the approximate problem $\mathcal{M}\tilde{\mathcal{SP}}^l$ is computationally tractable.

IV. NUMERICAL RESULTS

To evaluate the performance of the proposed decision rule approximations, we consider an instance of the electricity capacity expansion model discussed in [8]. The underlying power system consists of a set $R = \{1, \ldots, 5\}$ of regions with uncertain electricity demands δ_r, $r \in R$. Demands are satisfied by a set $N = \{1, 2, 3\}$ of power plans, where each plant $n \in N$ can produce up to $\bar{g}_n$ units of energy at uncertain costs ζ_n. Regions are connected by a set $M = \{1, \ldots, 5\}$ of directed transmission lines. Each line $m \in M$ has a capacity of $\bar{f}_m$ units of energy. The system topology is visualized in Figure 1.

The capacity expansion problem is modeled as the following two-stage stochastic program. In the first stage, we decide by how much the existing capacity of each plant $n \in N$ will be expanded at unit cost c_n and by how much the capacity of each transmission line $m \in M$ will be expanded at unit cost d_n. Then, the uncertain demands δ_r and operating costs ζ_n are revealed, which are assumed to be independent and uniformly distributed. In the second stage the expanded system is put into operation. The goal is to minimize the sum of investment costs and expected operating costs while satisfying all regional demands. We refer to [8] for more details.

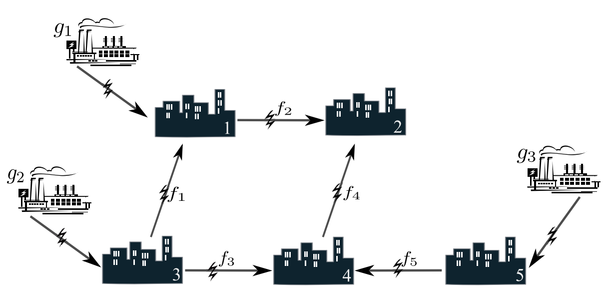

Fig. 1. Power System Configuration

The input parameters are summarized in Table I. We generate upper and lower bounds on the optimal value of the problem by using polynomial decision rules of various degrees and compute the relative optimality gaps. We compare these gaps with those obtained with existing methods based on linear [4] and piecewise linear [8] decision rules. The relative optimality gaps are computed by dividing the difference by the midpoint of the upper and lower bounds. All computations are performed within Matlab 2010b and using the Yalmip interface [15] of the SDPT3 optimization toolkit [16]. The employed piecewise linear decision rules have a general segmentation with 9 breakpoints per (primitive and composite) random parameter as described in [8]. The resulting relative optimality gaps for linear, piecewise linear, quadratic and cubic polynomial decision rules amount to 41%, 16% and 13% and 7%, respectively. Solving a capacity expansion problem of the type described here to within 7% accuracy is indeed sufficient for all practical purposes. The superior performance of polynomial decision rules with respect to linear and piecewise linear decision rules reflects their ability to adapt to different problem instances even if no structural information about the true optimal solution is available.

REFERENCES

[1] S.W. Wallace and W.T. Ziemba, *Applications of Stochastic Programming*, Society for Industrial and Applied Mathematics, 2005.

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

[3] A. Shapiro and A. Nemirovski, "On complexity of stochastic programming problems", *Continuous Optimization*, pp. 111–146, 2005.

[4] D. Kuhn, W. Wiesemann, and A. Georghiou, "Primal and Dual Linear Decision Rules in Stochastic and Robust Optimization", *Mathematical Programming*, pp. 1–33, 2009, available on Online First.

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

[6] X. Chen and Y. Zhang, "Uncertain linear programs: Extended affinely adjustable robust counterparts", *Operations Research*, vol. 57, no. 6, pp. 1469–1482, 2009.

[7] J. Goh and M. Sim, "Distributionally robust optimization and its tractable approximations", *Operations Research*, vol. 58, no. 4, pp. 902–917, 2010.

[8] A. Georghiou, W. Wiesemann, and D. Kuhn, "Generalized decision rule approximations for stochastic programming via liftings", *available on Optimization Online*, 2010.

[9] D. Bertsimas, D.A. Iancu, and P.A. Parrilo, "A hierarchy of near-optimal policies for multi-stage adaptive optimization", *Conditionally accepted in IEEE Transactions on Automatic Control*, 2009.

[10] M. Putinar, "Positive polynomials on compact semi-algebraic sets", *Indiana University Mathematics Journal*, vol. 42, no. 3, pp. 969–984, 1993.

[11] M. Dyer and L. Stougie, "Computational complexity of stochastic programming problems", *Mathematical Programming*, vol. 106, no. 3, pp. 423–432, 2006.

[12] Murty, "Some NP-complete problems in quadratic and nonlinear programming", *Mathematical Programming*, vol. 39, pp. 117–129, 1987.

[13] P.A. Parrilo, "Semidefinite programming relaxations for semialgebraic problems", *Mathematical Programming*, vol. 96, no. 2, pp. 293–320, 2003.

[14] E.K. Haviland, "On the momentum problem for distribution functions in more than one dimension", *American Journal of Mathematics*, vol. 57, no. 3, pp. 562–568, 1935.

[15] J. Löfberg, "YALMIP : A toolbox for modeling and optimization in MATLAB", in *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[16] R.H. Tütüncü, K.C. Toh, and M.J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3", *Mathematical Programming*, vol. 95, no. 2, pp. 189–217, 2003.