A Polynomial-Time Solution Scheme for Quadratic Stochastic Programs

Paula Rocha* Daniel Kuhn†

December 26, 2012

Abstract

We consider quadratic stochastic programs with random recourse — a class of problems which is perceived to be computationally demanding. Instead of using mainstream scenario tree-based techniques, we reduce computational complexity by restricting the space of recourse decisions to those linear and quadratic in the observations, thereby obtaining an upper bound on the original problem. To estimate the loss of accuracy of this approach, we further derive a lower bound by dualizing the original problem and solving it in linear and quadratic recourse decisions. By employing robust optimization techniques, we show that both bounding problems may be approximated by tractable conic programs.

Keywords: decision rule approximation, robust optimization, quadratic stochastic programming, conic programming

AMS Classification: 90C15

1 Introduction

We consider *quadratic* stochastic programs with *random recourse*. Despite their superior modeling power and frequent appearance in engineering and finance, these problems have received much less attention than standard linear stochastic programs with fixed recourse. The reasons for this negligence are as follows. The powerful L-shaped algorithm [1], that is, the most widely-used solution method for convex stochastic programs, only applies to problems with fixed recourse. Furthermore, the *recourse* or *cost-to-go*

*Corresponding author, Department of Computing, Imperial College London, London SW7 2AZ, United Kingdom, e-mail: paula.martins-da-silva-rocha08@imperial.ac.uk

†Department of Computing, Imperial College London, London SW7 2AZ, United Kingdom, e-mail: d.kuhn@imperial.ac.uk

functions of a multistage stochastic program with uncertainty-affected quadratic costs and/or random recourse are piecewise rational functions of the uncertain parameters and are, thus, not necessarily convex or concave. This complicates their (approximate) numerical integration and the estimation of the corresponding approximation error. Indeed, stochastic programs with random recourse have resisted quantitative stability analysis until very recently [2].

In this paper, we propose a new solution scheme for quadratic stochastic programs with random recourse. Instead of approximating the stochastic process of the uncertain parameters by a scenario tree (see, e.g., [3]), we approximate the recourse decisions or *decision rules* by linear or quadratic functions of the uncertain parameters. This restriction of the decision maker's flexibility results in an *upper* bound on the true optimal value of the stochastic program. Conversely, by solving the dual of the original stochastic program in linear and/or quadratic decision rules, we obtain a *lower* bound. We demonstrate that both bounding problems can be conservatively approximated by tractable (i.e., polynomial-time solvable) conic programs. The gap between their optimal values estimates the loss of optimality incurred by the decision rule approximation. Since both conic programs scale polynomially with the size of the problem description, our approach is expected to unfold its full potential when applied to large-scale problems with many decision stages.

Fueled by the recent progress in modern convex optimization, decision rule techniques of the type proposed here have found successful application in worst-case robust optimization [4], distributionally robust optimization [5] and stochastic programming [6]. The main focus of previous work has been on *primal linear* decision rules for *linear* stochastic and robust optimization problems with *fixed* recourse. Only few authors have studied *piecewise linear* [5, 7, 8] or *polynomial* [9, 10] decision rules. Lower bounds based on *dual* decision rule approximations were first discussed in [11].

The key contributions of this paper can be summarized as follows. We develop an efficient decision rule approximation for quadratic stochastic programs with random recourse. While the genuine decision variables are approximated by linear decision rules, the stochasticity of the constraint matrices and the Hessian of the objective function prompts us to model the analysis and slack variables as quadratic decision rules. Moreover, we propose a systematic method for estimating the degree of suboptimality of the best linear-quadratic decision rule. Our approach differs substantially from the method described in [11] for *linear* multistage models with *fixed* recourse — a method which cannot be extended to quadratic models with random recourse. Indeed, the new approach presented here is not based on a constraint aggregation emerging from an *implicit* dualization of the original stochastic program, but arises from an *explicit* dualization and subsequent decision rule approximation.

2 Quadratic One-Stage Stochastic Program with Random Recourse

The following notation is used in the remainder of the paper. Uncertainty is modeled by a probability space $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mathbb{P})$. The Borel σ-algebra $\mathcal{B}(\mathbb{R}^k)$ is the set of events that are assigned probabilities by the probability measure $\mathbb{P}$. We denote by ξ the elements of the sample space $\mathbb{R}^k$ and by Ξ the support of $\mathbb{P}$, i.e., the smallest closed subset of $\mathbb{R}^k$ which has probability 1. $\mathbb{E}(\cdot)$ denotes the expectation operator with respect to $\mathbb{P}$. We let $\mathcal{L}_{k,n}$ represent the space of all measurable functions from $\mathbb{R}^k$ to $\mathbb{R}^n$ that are bounded on compact sets. By slight abuse of notation, for $A, B \in \mathbb{R}^{m\times n}$ the relation $A \geq B$ represents componentwise inequality. For $C, D \in \mathbb{R}^{n\times n}$, the relation $C \succeq D$ implies that $C - D$ is positive semidefinite. We denote the Euclidean norm in $\mathbb{R}^n$ by $\|\cdot\|_2$. The second-order cone in $\mathbb{R}^{n+1}$ is $\mathcal{K}_2 := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : \|x\|_2 \leq t\}$. For any given proper cone $\mathcal{K}$, the relation $y \succeq_{\mathcal{K}} z$ implies that $(y - z) \in \mathcal{K}$. The converse inequalities $C \leq D$, $C \preceq D$ and $y \preceq_{\mathcal{K}} z$ are defined in the obvious way. For every $C \in \mathbb{R}^{n\times n}$, we let $\mathrm{tr}(C)$ denote the trace of C. Moreover, the operator $\otimes$ stands for the Kronecker product. We denote by $\mathbb{S}^k$ the space of symmetric matrices in $\mathbb{R}^{k\times k}$. Finally, we denote by e_n the n-th canonical basis vector. Its dimension will normally be clear from the context.

We consider decision problems under uncertainty of the following type. A decision maker observes an element ξ of the sample space $\mathbb{R}^k$ and then chooses a decision $x(\xi) \in \mathbb{R}^n$ subject to the constraints $A(\xi)x(\xi) \leq b(\xi)$ and $x(\xi) \geq 0$. The decision rule $x \in \mathcal{L}_{k,n}$ is selected in such a way so as to minimize the expected value of $\frac{1}{2}x(\xi)^\top Q(\xi)x(\xi) + c(\xi)^\top x(\xi)$. This decision problem can be formulated as the following quadratic one-stage stochastic program:

$$\mathcal{P}^o: \quad \begin{array}{ll} \inf & \mathbb{E}\Big(\frac{1}{2}x(\xi)^\top Q(\xi)x(\xi) + c(\xi)^\top x(\xi)\Big) \\ \text{s.t.} & x \in \mathcal{L}_{k,n} \\ & \left.\begin{array}{l} A(\xi)x(\xi) \leq b(\xi) \\ x(\xi) \geq 0 \end{array}\right\} \mathbb{P}\text{-a.s.} \end{array}$$

To guarantee that $\mathcal{P}^o$ is well-defined, we require that the underlying problem data satisfies the following conditions. We first assume that $Q(\xi) \in \mathbb{R}^{n\times n}$ is symmetric and positive semidefinite with rank r. Under this assumption, there exists a full column rank matrix $F(\xi) \in \mathbb{R}^{n\times r}$ such that $Q(\xi) = F(\xi)F(\xi)^\top$. The cost coefficients $c(\xi)$, the right-hand side vector $b(\xi)$, the recourse matrix $A(\xi)$ and the matrix $F(\xi)$ are assumed to depend linearly on the random data. Formally speaking, we postulate that $c(\xi) = C\xi$ for some matrix $C \in \mathbb{R}^{n\times k}$ and $b(\xi) = B\xi$ for some matrix $B \in \mathbb{R}^{m\times k}$. Moreover, the μ-th row of

$A(\xi)$ is representable as $\tilde{a}_\mu(\xi)^\top = \xi^\top \tilde{A}_\mu$ for some matrix $\tilde{A}_\mu \in \mathbb{R}^{k\times n}$, where μ ranges from 1 to m. This implies that the ν-th column of $A(\xi)$ may be written as $a_\nu(\xi) = A_\nu \xi$ for $\nu = 1,\dots,n$, where $A_\nu := (\tilde{A}_1 e_\nu, \dots, \tilde{A}_m e_\nu)^\top$. Finally, the ν-th row of $F(\xi)$ may be expressed as $\tilde{f}_\nu(\xi)^\top = \xi^\top \tilde{F}_\nu$ for some matrix $\tilde{F}_\nu \in \mathbb{R}^{k\times r}$, where ν ranges from 1 to n. Therefore, the ρ-th column of $F(\xi)$ is representable as $f_\rho(\xi) = F_\rho \xi$, where ρ ranges from 1 to r and $F_\rho := (\tilde{F}_1 e_\rho, \dots, \tilde{F}_n e_\rho)^\top$. Note that these linearity assumptions are non-restrictive since we may redefine the vector ξ as the concatenation of all components of $c(\xi)$, $b(\xi)$, $A(\xi)$ and $F(\xi)$, if necessary.

We further require the support of $\mathbb{P}$ to be a non-empty and compact set of the form

$$\Xi = \{\xi \in \mathbb{R}^k : e_1^\top \xi = 1, \xi^\top O_\ell \xi \geq 0, \ell = 1, \dots, l\}, \tag{1}$$

where $O_\ell \in \mathbb{S}^k$ is representable as

$$O_\ell = \begin{pmatrix} \omega\ell & o_\ell^\top \\ o_\ell & -\Omega_\ell \Omega_\ell^\top \end{pmatrix}$$

for some $\Omega_\ell \in \mathbb{R}^{(k-1)\times q_\ell}$, $o_\ell \in \mathbb{R}^{k-1}$ and $\omega_\ell \in \mathbb{R}$. By construction, the first component of every $\xi \in \Xi$ is equal to 1. This specification allows us to represent affine functions of the nondegenerate outcomes $(\xi_2, \dots, \xi_k)^\top$ concisely as linear functions of $\xi := (\xi_1, \dots, \xi_k)^\top$. It also enables us to represent every quadratic function in $(\xi_2, \dots, \xi_k)^\top$ as a homogeneous function of degree 2 in ξ. We further assume that Ξ spans the whole sample space $\mathbb{R}^k$. This is true iff the system $\xi^\top O_\ell \xi \geq 0, \ell = 1, \dots, l$, is strictly feasible.

Remark 2.1. *All compact subsets of the hyperplane* $\{\xi \in \mathbb{R}^k : e_1^\top \xi = 1\}$ *that result from intersections of closed halfspaces and ellipsoids can be represented as sets of the form* (1)*.*

For further argumentation, it proves useful to introduce new decision rules $z \in \mathcal{L}_{k,m}$ and $y \in \mathcal{L}_{k,r}$ in $\mathcal{P}^o$ to convert the first inequality into an equality constraint and to eliminate $Q(\xi)$ from the objective function, respectively. Thus, $\mathcal{P}^o$ can be equivalently expressed as

$$\mathcal{P}: \quad \begin{array}{ll} \inf & \mathbb{E}\Big(\frac{1}{2} y(\xi)^\top y(\xi) + c(\xi)^\top x(\xi)\Big) \\ \text{s.t.} & x \in \mathcal{L}_{k,n}, y \in \mathcal{L}_{k,r}, z \in \mathcal{L}_{k,m} \\ & \left.\begin{array}{l} A(\xi)x(\xi) + z(\xi) = b(\xi) \\ y(\xi) = F(\xi)^\top x(\xi) \\ x(\xi) \geq 0, z(\xi) \geq 0 \end{array}\right\} \mathbb{P}\text{-a.s.} \end{array}$$

2.1 Primal Approximation

Problem $\mathcal{P}$ is computationally intractable since it constitutes an optimization problem over an infinite-dimensional function space. To facilitate numerical tractability, we restrict the functional form of $x(\xi)$ to be linear in the uncertain parameters, that is, we require that $x(\xi) = X\xi$ for some matrix $X \in \mathbb{R}^{n\times k}$. As a result, the product terms $A(\xi)x(\xi)$ and $F(\xi)^\top x(\xi)$ become quadratic functions of ξ. Therefore, the equality constraints are only satisfiable if the decisions $z(\xi)$ and $y(\xi)$ exhibit a quadratic dependence on the random data. Thus, we require that $z_\mu(\xi) = \xi^\top Z_\mu \xi$ and $y_\rho(\xi) = \xi^\top Y_\rho \xi$ for some (without any loss of generality symmetric) matrices $Z_\mu, Y_\rho \in \mathbb{S}^k$, where $\mu = 1,\dots,m$ and $\rho = 1,\dots,r$. With these conventions, problem $\mathcal{P}$ reduces to

$$
\mathcal{P}^u: \quad
\begin{array}{ll}
\inf & \displaystyle\frac{1}{2}\sum_{\rho=1}^{r} \operatorname{tr}\Big(\mathbb{E}(\xi\xi^\top \otimes \xi\xi^\top)(Y_\rho \otimes Y_\rho)\Big) + \operatorname{tr}\Big(\mathbb{E}(\xi\xi^\top) C^\top X\Big) \\
\text{s.t.} & X \in \mathbb{R}^{n\times k}, Y_1,\dots,Y_r, Z_1,\dots,Z_m \in \mathbb{S}^k \\
& \left.\begin{array}{l}
\xi^\top \tilde{A}_\mu X\xi + \xi^\top Z_\mu \xi = \tilde{b}_\mu^\top \xi \quad \forall \mu = 1,\dots,m \\
\xi^\top Y_\rho \xi = \xi^\top F_\rho^\top X\xi \quad \forall \rho = 1,\dots,r \\
X\xi \geq 0 \\
\xi^\top Z_\mu \xi \geq 0 \quad \forall \mu = 1,\dots,m
\end{array}\right\} \mathbb{P}\text{-a.s.,}
\end{array}
$$

where $\tilde{b}_\mu^\top$ denotes the μ-th row of the matrix B. Since problem $\mathcal{P}^u$ was obtained by restricting the underlying feasible set, it provides an upper bound on $\mathcal{P}$. Even though the number of decision variables is now finite, $\mathcal{P}^u$ still appears to be intractable since it involves infinitely many constraints. However, using techniques from modern robust optimization [12], we can demonstrate that $\mathcal{P}^u$ is, in fact, tractable. We begin by observing that, due to their continuity in ξ, the almost sure constraints in $\mathcal{P}^u$ hold for all $\xi \in \Xi$. Thus, the μ-th equality constraint is equivalent to

$$
\xi^\top H_\mu \xi = 0 \quad \forall \xi \in \Xi, \tag{2}
$$

where $H_\mu \in \mathbb{S}^k$ is defined as

$$
H_\mu := \frac{1}{2}\big(\tilde{A}_\mu X + X^\top \tilde{A}_\mu^\top - e_1 \tilde{b}_\mu^\top - \tilde{b}_\mu e_1^\top\big) + Z_\mu.
$$

By its homogeneity, (2) extends to cone(Ξ), i.e., the cone generated by Ξ. Thus, the Hessian of the mapping $\xi \mapsto \xi^\top H_\mu \xi$, which is given by $2H_\mu$, vanishes in the interior of cone(Ξ). As Ξ spans $\mathbb{R}^k$, the interior of cone(Ξ) is non-empty. Therefore, we conclude that $H_\mu = 0$, so the first set of equality constraints in $\mathcal{P}^u$ is equivalent to the requirement that $H_\mu = 0$ for all $\mu = 1, \ldots, m$. A similar argument applies to the second set of equality constraints. Simplification of the semi-infinite constraint $X\xi \geq 0$ $\mathbb{P}$-a.s. relies on the following proposition, which can be regarded as a special case of a central result in robust optimization; see, e.g., [4, Theorem 3.2].

Proposition 2.1. *Consider the following two convex cones in* $\mathbb{R}^k$*:*

$$\begin{aligned}
\mathcal{K} &:= \left\{z \in \mathbb{R}^k : z^\top \xi \geq 0 \quad \forall \xi \in \Xi\right\}, \\
\hat{\mathcal{K}} &:= \Big\{z \in \mathbb{R}^k : \exists \psi \in \mathbb{R}, \phi_\ell \in \mathbb{R}^{q_\ell}, \ell = 1, \ldots, l, \text{ with } z = e_1 \psi + \textstyle\sum_{\ell=1}^{l} \hat{O}_\ell^\top \phi_\ell, \\
&\qquad \psi \geq 0 \text{ and } \phi_\ell \succeq_{\mathcal{K}_2} 0\Big\},
\end{aligned}$$

where

$$\hat{O}_\ell := \begin{bmatrix} 0 & \Omega_\ell^\top \\ \frac{1-\omega_\ell}{2} & -o_\ell^\top \\ \frac{1+\omega_\ell}{2} & o_\ell^\top \end{bmatrix}.$$

Then, $\mathcal{K} = \hat{\mathcal{K}}$*.*

Note that the first inequality constraint in $\mathcal{P}^u$ is equivalent to the requirement that every row of X belongs to the cone $\mathcal{K}$. Using Proposition 2.1, we may re-express this constraint as $X \in \hat{\mathcal{K}}^n$, where we interpret the Cartesian product $\hat{\mathcal{K}}^n$ as the cone of all $n \times k$ matrices whose rows are contained in $\hat{\mathcal{K}}$. Lastly, to approximate the semi-infinite constraints $\xi^\top Z_\mu \xi \geq 0$ $\mathbb{P}$-a.s., $\mu = 1, \ldots, m$, we apply the following proposition.

Proposition 2.2. *Consider the following two convex cones in* $\mathbb{S}^k$*:*

$$\begin{aligned}
\mathcal{C} &:= \left\{S \in \mathbb{S}^k : \xi^\top S \xi \geq 0 \quad \forall \xi \in \Xi\right\}, \\
\hat{\mathcal{C}} &:= \left\{S \in \mathbb{S}^k : \exists \lambda \in \mathbb{R}^l \text{ with } \lambda \geq 0 \text{ and } S - \textstyle\sum_{\ell=1}^{l} \lambda_\ell O_\ell \succeq 0\right\}.
\end{aligned}$$

Then, $\hat{\mathcal{C}} \subseteq \mathcal{C}$ *for each* $l \in \mathbb{N}$*, and* $\hat{\mathcal{C}} = \mathcal{C}$ *if* $l = 1$*.*

The assertions in Proposition 2.2 follow from the approximate and exact versions of the S-Lemma, respectively (see, e.g., [4, Section 4] or [11, Proposition 6]). The last set of constraints in $\mathcal{P}^u$ is equivalent to the requirement that $Z_\mu \in \mathcal{C}$ for each $\mu = 1, \ldots, m$. Restricting these constraints to $Z_\mu \in \hat{\mathcal{C}}, \mu = 1, \ldots, m$, yields the following convex conic optimization problem:

$$
\begin{aligned}
\hat{\mathcal{P}}^u: \qquad & \inf && \frac{1}{2}\sum_{\rho=1}^{r} \operatorname{tr}\left(\mathbb{E}(\xi\xi^\top \otimes \xi\xi^\top)(Y_\rho \otimes Y_\rho)\right) + \operatorname{tr}\left(\mathbb{E}(\xi\xi^\top)C^\top X\right) \\
& \text{s.t.} && X \in \mathbb{R}^{n\times k}, Y_1, \ldots, Y_r, Z_1, \ldots, Z_m \in \mathbb{S}^k \\
& && \tfrac{1}{2}\left(\tilde{A}_\mu X + X^\top \tilde{A}_\mu^\top\right) + Z_\mu = \tfrac{1}{2}\left(e_1 \tilde{b}_\mu^\top + \tilde{b}_\mu e_1^\top\right) \quad \forall \mu = 1, \ldots, m \\
& && \tfrac{1}{2}\left(F_\rho^\top X + X^\top F_\rho\right) - Y_\rho = 0 \quad \forall \rho = 1, \ldots, r \\
& && X \succeq_{\hat{\mathcal{K}}^n} 0 \\
& && Z_\mu \succeq_{\hat{\mathcal{C}}} 0 \quad \forall \mu = 1, \ldots, m.
\end{aligned}
$$

The above reasoning implies that the conic program $\hat{\mathcal{P}}^u$ provides a conservative approximation (when $l > 1$) or an exact reformulation (when $l = 1$) for $\mathcal{P}^u$. By using the definitions of $\hat{\mathcal{K}}^n$ and $\hat{\mathcal{C}}$ to expand the conic constraints, problem $\hat{\mathcal{P}}^u$ can be reformulated as an explicit semidefinite program (SDP), whose size is polynomial in k, l, m and n. Therefore, it is amenable to efficient numerical solution via modern interior-point algorithms [13]. We remark that $\hat{\mathcal{P}}^u$ only requires information about the support and the moments (up to fourth-order) of the uncertain parameters — an attractive feature from a modeling perspective since the full joint probability distribution of ξ is seldom available.

Remark 2.2. *If problem $\mathcal{P}^o$ has fixed recourse, that is, if $A(\xi) = A$ for some $A \in \mathbb{R}^{m\times n}$, then Z_μ is representable as*

$$
Z_\mu = \begin{pmatrix} \zeta_\mu & \frac{1}{2} z_\mu^\top \\ \frac{1}{2} z_\mu & 0 \end{pmatrix}
$$

for some $\zeta_\mu \in \mathbb{R}$ and $z_\mu \in \mathbb{R}^{k-1}, \mu = 1, \ldots, m$. Then, the condition $Z_\mu \in \mathcal{C}$ is equivalent to $Z_\mu \in \hat{\mathcal{C}}$, see [14, *Proposition* 3.7]*, and may ultimately be simplified to $(\zeta_\mu, z_\mu^\top)^\top \in \hat{\mathcal{K}}$ using techniques described in* [15, 16]*. We conclude that, for fixed recourse problems, $\mathcal{P}^u$ is equivalent to a second-order cone program.*

2.2 Dual Problem

To estimate the loss of optimality incurred by the decision rule approximation proposed in Section 2.1, we now determine a computationally tractable lower bound on $\mathcal{P}^o$. To this end, we dualize $\mathcal{P}$, apply a decision rule approximation to the dual problem and simplify the resulting problem by using robust optimization techniques. In the following, we let '$\inf_{x,y,z}$' be a shorthand notation for the infimum operator over all $x \in \mathcal{L}_{k,n}$, $y \in \mathcal{L}_{k,r}$ and $z \in \mathcal{L}_{k,m}$. Similarly, we let '$\sup_{s,u,v,w}$' denote the supremum operator over all $u \in \mathcal{L}_{k,m}$ and $v \in \mathcal{L}_{k,r}$ as well as over all $s \in \mathcal{L}_{k,n}$ and $w \in \mathcal{L}_{k,m}$ that are almost surely non-negative. By assigning the dual decision rules (i) $u \in \mathcal{L}_{k,m}$, (ii) $v \in \mathcal{L}_{k,r}$, (iii) $s \in \mathcal{L}_{k,n}$ and (iv) $w \in \mathcal{L}_{k,m}$ to the constraints (i) $A(\xi)x(\xi) + z(\xi) = b(\xi)$ $\mathbb{P}$-a.s., (ii) $y(\xi) = F(\xi)^\top x(\xi)$ $\mathbb{P}$-a.s., (iii) $x(\xi) \geq 0$ $\mathbb{P}$-a.s. and (iv) $z(\xi) \geq 0$ $\mathbb{P}$-a.s., respectively, we build the following Lagrangian for problem $\mathcal{P}$:

$$\begin{aligned} L(x, y, z; s, u, v, w) \quad := \quad & \mathbb{E}\Big\{ \tfrac{1}{2} y(\xi)^\top y(\xi) + c(\xi)^\top x(\xi) + u(\xi)^\top \left[A(\xi)x(\xi) + z(\xi) - b(\xi)\right] + \\ & v(\xi)^\top \left[F(\xi)^\top x(\xi) - y(\xi)\right] - s(\xi)^\top x(\xi) - w(\xi)^\top z(\xi) \Big\}. \end{aligned} \tag{3}$$

Using this Lagrangian, problem $\mathcal{P}$ may be reformulated as

$$\inf_{x,y,z} \sup_{s,u,v,w} L(x, y, z; s, u, v, w), \tag{4}$$

while its corresponding dual problem is defined as

$$\sup_{s,u,v,w} \inf_{x,y,z} L(x, y, z; s, u, v, w). \tag{5}$$

Problems $\mathcal{P}$ and (4) are equivalent since the inner maximization over the dual decision rules in (4) imposes an infinite penalty on any primal decision $(x, y, z) \in \mathcal{L}_{k,n} \times \mathcal{L}_{k,r} \times \mathcal{L}_{k,m}$ which violates the almost sure constraints in $\mathcal{P}$ on a set of strictly positive probability. We remark that the equivalence between problems $\mathcal{P}$ and (4) holds even if $\mathcal{P}$ is infeasible. For a formal proof of this equivalence, we refer to [17, Section 4]. By weak duality, the supremum in (5) provides a lower bound on the infimum in (4); see, e.g., [17, Theorem 4]. This is true also when the supremum in (5) or the infimum in (4) are infinite.

A set of optimality conditions for the inner minimization problem in (5) is found by setting the Gâteaux differential of the Lagrangian (3) with respect to x, y and z to zero, for all descent directions $h_\mathrm{x} \in \mathcal{L}_{k,n}$, $h_\mathrm{y} \in \mathcal{L}_{k,r}$ and $h_\mathrm{z} \in \mathcal{L}_{k,m}$, respectively; see, e.g., [18, Section 7.2]:

$$\begin{aligned} \mathbb{E}\Big\{ h_\mathrm{x}(\xi)^\top \left[c(\xi) + A(\xi)^\top u(\xi) + F(\xi)v(\xi) - s(\xi)\right] \Big\} &= 0 \quad \forall h_\mathrm{x} \in \mathcal{L}_{k,n} \\ \mathbb{E}\Big\{ h_\mathrm{y}(\xi)^\top \left[y(\xi) - v(\xi)\right] \Big\} &= 0 \quad \forall h_\mathrm{y} \in \mathcal{L}_{k,r} \\ \mathbb{E}\Big\{ h_\mathrm{z}(\xi)^\top \left[u(\xi) - w(\xi)\right] \Big\} &= 0 \quad \forall h_\mathrm{z} \in \mathcal{L}_{k,m}. \end{aligned}$$

These conditions are equivalent to

$$A(\xi)^\top u(\xi) + F(\xi)v(\xi) - s(\xi) = -c(\xi) \quad \mathbb{P}\text{-a.s.} \tag{6}$$

$$y(\xi) = v(\xi) \quad \mathbb{P}\text{-a.s.} \tag{7}$$

$$w(\xi) = u(\xi) \quad \mathbb{P}\text{-a.s.} \tag{8}$$

Thus, we find that

$$\inf_{x,y,z} L(x,y,z;s,u,v,w) = \begin{cases} -\mathbb{E}\Big(\frac{1}{2}v(\xi)^\top v(\xi) + b(\xi)^\top u(\xi)\Big), & \text{if (6) and (8) hold} \\ -\infty, & \text{otherwise.} \end{cases} \tag{9}$$

Substituting (9) into (5) yields the following concave quadratic stochastic program dual to $\mathcal{P}$:

$$\mathcal{D}: \quad \begin{array}{ll} \sup & -\mathbb{E}\Big(\frac{1}{2}v(\xi)^\top v(\xi) + b(\xi)^\top u(\xi)\Big) \\ \text{s.t.} & u \in \mathcal{L}_{k,m}, v \in \mathcal{L}_{k,r}, s \in \mathcal{L}_{k,n} \\ & \left.\begin{array}{l} A(\xi)^\top u(\xi) + F(\xi)v(\xi) - s(\xi) = -c(\xi) \\ u(\xi) \geq 0, s(\xi) \geq 0 \end{array}\right\} \mathbb{P}\text{-a.s.} \end{array}$$

Like its primal counterpart, problem $\mathcal{D}$ is computationally intractable as it involves a continuum of decision variables and constraints. As the problems $\mathcal{P}$ and $\mathcal{D}$ have essentially the same structure, we can proceed as in Section 2.1 to derive a tractable approximation for $\mathcal{D}$. For brevity, we will omit some details of this derivation.

2.3 Dual Approximation

To reduce the complexity of problem $\mathcal{D}$, we restrict the functional form of the dual decision rules to those which may be represented as

$$u(\xi) = U\xi, \quad v(\xi) = V\xi \quad \text{and} \quad s_\nu(\xi) = \xi^\top S_\nu \xi$$

for some matrices $U \in \mathbb{R}^{m\times k}$, $V \in \mathbb{R}^{r\times k}$ and $S_\nu \in \mathbb{S}^k, \nu = 1,\dots,n$. Using this decision rule approximation, problem $\mathcal{D}$ simplifies to

$$\mathcal{D}^l: \quad \begin{array}{ll} \sup & -\frac{1}{2}\operatorname{tr}\left\{V\mathbb{E}(\xi\xi^\top)V^\top\right\} - \operatorname{tr}\left\{\mathbb{E}(\xi\xi^\top)B^\top U\right\} \\ \text{s.t.} & U \in \mathbb{R}^{m\times k}, V \in \mathbb{R}^{r\times k}, S_1,\dots,S_n \in \mathbb{S}^k \\ & \left.\begin{array}{l} \xi^\top A_\nu^\top U\xi + \xi^\top \tilde{F}_\nu V\xi - \xi^\top S_\nu \xi = -\tilde{c}_\nu^\top \xi \quad \forall \nu = 1,\dots,n \\ U\xi \geq 0 \\ \xi^\top S_\nu \xi \geq 0 \quad \forall \nu = 1,\dots,n \end{array}\right\} \mathbb{P}\text{-a.s.,} \end{array}$$

where $\tilde{c}_\nu^\top$ denotes the ν-th row of the matrix C. Problem $\mathcal{D}^l$ provides a lower bound on $\mathcal{D}$ as it was obtained by reducing the underlying feasible set. Employing the same robust optimization techniques as in Section 2.1, $\mathcal{D}^l$ can be approximated by the following conic program:

$$
\begin{aligned}
\hat{\mathcal{D}}^l : \qquad & \sup \quad -\tfrac{1}{2}\operatorname{tr}\left\{V\mathbb{E}(\xi\xi^\top)V^\top\right\} - \operatorname{tr}\left\{\mathbb{E}(\xi\xi^\top)B^\top U\right\} \\
& \text{s.t.} \quad U \in \mathbb{R}^{m\times k}, V \in \mathbb{R}^{r\times k}, S_1,\dots,S_n \in \mathbb{S}^k \\
& \qquad \tfrac{1}{2}\left(A_\nu^\top U + U^\top A_\nu + \tilde{F}_\nu V + V^\top \tilde{F}_\nu^\top\right) - S_\nu = -\tfrac{1}{2}\left(e_1\tilde{c}_\nu^\top + \tilde{c}_\nu e_1^\top\right) \quad \forall \nu = 1,\dots,n \\
& \qquad U \succeq_{\hat{\mathcal{K}}^m} 0 \\
& \qquad S_\nu \succeq_{\hat{\mathcal{C}}} 0 \quad \forall \nu = 1,\dots,n.
\end{aligned}
$$

From Proposition 2.2, we conclude that $\hat{\mathcal{D}}^l$ constitutes a conservative approximation (when $l > 1$) or an exact reformulation (when $l = 1$) for $\mathcal{D}^l$. Like its primal counterpart $\hat{\mathcal{P}}^u$, problem $\hat{\mathcal{D}}^l$ can be explicitly expressed as an SDP by expanding the conic constraints. The size of this SDP is polynomial in k, l, m and n, implying that it can be solved efficiently. Note that the probability distribution of ξ only affects $\hat{\mathcal{D}}^l$ through its first- and second-order moments and its support.

Remark 2.3. *If $A(\xi) = A$ and $F(\xi) = F$ for some $A \in \mathbb{R}^{m\times n}$ and $F \in \mathbb{R}^{n\times r}$, then problem $\mathcal{D}^l$ can be reformulated as a second-order cone program. If either (i) $A(\xi) = A$ or (ii) $F(\xi) = F$, then one can improve the approximation quality by modeling (i) $u(\xi)$ or (ii) $v(\xi)$ as quadratic decision rules, respectively. This refinement preserves the complexity class of problem $\hat{\mathcal{D}}^l$, which remains an SDP.*

A summary of the key insights of this section is provided in the following Theorem.

Theorem 2.1. *The following relation holds:*

$$
\sup \hat{\mathcal{D}}^l \le \sup \mathcal{D}^l \le \sup \mathcal{D} \le \inf \mathcal{P} \le \inf \mathcal{P}^u \le \inf \hat{\mathcal{P}}^u,
$$

where the first and last inequalities convert into equalities if $l = 1$. Problems $\hat{\mathcal{D}}^l$ and $\hat{\mathcal{P}}^u$ can be solved in polynomial time. The gap between their optimal values provides an estimate of the loss of accuracy incurred by the adopted decision rule approximation.

3 Quadratic Multi-Stage Stochastic Program with Random Recourse

In this section, we continue to assume that $\mathbb{P}$ has a polyhedral support Ξ of the type (1) that is non-empty, bounded, and spans $\mathbb{R}^k$. Now, however, we impose a temporal structure on the elements of the sample space. More concretely, we assume that ξ can be partitioned into subvectors of the form $(\xi_{k^{t-1}+1}, \dots, \xi_{k^t})$ for some $k^0 = 0$ and $1 = k^1 < k^2 < \dots < k^T = k$, which are observed sequentially at times $t \in \mathcal{T} := \{1, \dots, T\}$, respectively. We denote the history of observations up to time t by $\xi^t := (\xi_1, \dots, \xi_{k^t}) \in \mathbb{R}^{k^t}$. By construction, $\xi^T = \xi$. Moreover, we let $\mathbb{E}_t(\cdot)$ denote the expectation with respect to $\mathbb{P}$ conditional on ξ^t.

We consider a sequential decision process in which the decision $x_\nu(\xi^t) \in \mathbb{R}$, $\nu \in \mathcal{N}_t := \{n^{t-1}+1, \dots, n^t\}$, is selected at time t after observing ξ^t, but before the future outcomes $\xi_{k^t+1}, \dots, \xi_k$ are known. Here, it is understood that $0 = n^0 < n^1 < \dots < n^T = n$, and we set $n_t := n^t - n^{t-1}$. The aim is to find a sequence of *non-anticipative* decision rules $x := (x_1, \dots, x_n) \in \mathcal{G} := \times_{t=1}^T \mathcal{L}_{k^t, n_t}$ which map the available observations to decisions while minimizing a quadratic expected cost function subject to linear constraints. Such decision problems may be formulated as quadratic multistage stochastic programs of the form

$$\mathcal{MP}^o: \qquad \begin{array}{ll} \inf & \mathbb{E}\Big(\tfrac{1}{2}x(\xi)^\top Q(\xi) x(\xi) + c(\xi)^\top x(\xi)\Big) \\ \text{s.t.} & x \in \mathcal{G} \\ & \left.\begin{array}{l} \mathbb{E}_t\Big(\tilde{a}_\mu(\xi)^\top x(\xi)\Big) \le b_\mu(\xi^t) \quad \forall \mu \in \mathcal{M}_t, t \in \mathcal{T} \\ x(\xi) \ge 0 \end{array}\right\} \mathbb{P}\text{-a.s.,} \end{array}$$

where $\mathcal{M}_t := \{m^{t-1}+1, \dots, m^t\}$ for some $0 = m^0 < m^1 < \dots < m^T = m$. For $\mathcal{MP}^o$ to be well-defined, $Q(\xi)$ is assumed to be symmetric and positive semidefinite with rank r, and, thus, it may be factorized as $Q(\xi) = F(\xi)F(\xi)^\top$ for some $F(\xi) \in \mathbb{R}^{n \times r}$. Furthermore, the linearity assumptions on $c(\xi)$, $b(\xi)$, $A(\xi)$ and $F(\xi)$ described in Section 2 still hold. In addition, we postulate that the cost coefficients $c(\xi)$ and the right-hand side vector $b(\xi)$ can be written as non-anticipative linear functions of the random parameters. Therefore, the matrices C and B are assumed to be representable as

$$C = \begin{bmatrix} C_1 P_1 \\ \vdots \\ C_T P_T \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} B_1 P_1 \\ \vdots \\ B_T P_T \end{bmatrix}$$

for some matrices $C_t \in \mathbb{R}^{n_t \times k^t}$ and $B_t \in \mathbb{R}^{m_t \times k^t}, t \in \mathcal{T}$, where $m_t := m^t - m^{t-1}$. Here, we used the truncation operators $P_t, t \in \mathcal{T}$, defined through $P_t : \mathbb{R}^k \mapsto \mathbb{R}^{k^t}, \xi \mapsto \xi^t$. By introducing the decision rules $y \in \mathcal{L}_{k,r}$ and $z := (z_1, \dots, z_m) \in \mathcal{H} := \times_{t=1}^T \mathcal{L}_{k^t, m_t}$, problem $\mathcal{MP}^o$ can be converted to

$$\begin{aligned}
\mathcal{MP}: \quad & \inf \quad \mathbb{E}\Big(\tfrac{1}{2}y(\xi)^\top y(\xi) + c(\xi)^\top x(\xi)\Big) \\
& \text{s.t.} \quad x \in \mathcal{G}, y \in \mathcal{L}_{k,r}, z \in \mathcal{H} \\
& \qquad \left.\begin{aligned}
& \mathbb{E}_t\Big(\tilde{a}_\mu(\xi)^\top x(\xi)\Big) + z_\mu(\xi^t) = b_\mu(\xi^t) \quad \forall \mu \in \mathcal{M}_t, t \in \mathcal{T} \\
& y(\xi) = F(\xi)^\top x(\xi) \\
& x(\xi) \geq 0, z(\xi) \geq 0
\end{aligned}\right\} \mathbb{P}\text{-a.s.}
\end{aligned}$$

Note that problems $\mathcal{MP}$ and $\mathcal{P}$ have a very similar structure. Thus, we may follow the same general strategy as in Section 2 to derive tractable bounding problems. For the sake of compactness, we will abbreviate the involved derivations.

3.1 Primal Approximation

In an attempt to reduce computational complexity of problem $\mathcal{MP}$, we restrict our attention to primal decision rules that are representable as

$$x(\xi) = X\xi, \quad y_\rho(\xi) = \xi^\top Y_\rho \xi \quad \text{and} \quad z_\mu(\xi^t) = (\xi^t)^\top Z_\mu \xi^t \tag{10}$$

for some matrices $Y_\rho \in \mathbb{S}^k$ and $Z_\mu \in \mathbb{S}^{k^t}$, where $\rho = 1, \dots, r$ and $\mu \in \mathcal{M}_t, t \in \mathcal{T}$. Here, the matrix X is assumed to belong to the linear space $\mathcal{X}$ of all block triangular matrices of the form

$$X = \begin{bmatrix} X_1 P_1 \\ \vdots \\ X_T P_T \end{bmatrix}$$

for some $X_t \in \mathbb{R}^{n_t \times k^t}, t \in \mathcal{T}$. To ensure that this approximation will convert $\mathcal{MP}$ to a tractable problem, we require $\mathbb{E}_t(\xi\xi^\top)$ to be almost surely quadratic in ξ^t. Formally speaking, we postulate that there exists a matrix $\Sigma_{t\kappa\kappa'} \in \mathbb{S}^{k^t}$ such that almost surely $\mathbb{E}_t(\xi_\kappa \xi_{\kappa'}) = (\xi^t)^\top \Sigma_{t\kappa\kappa'} \xi^t$, for each $t \in \mathcal{T}$ and $\kappa, \kappa' = 1, \dots, k$. This condition is trivially satisfied if, for instance, the random parameters are stagewise independent. By solving $\mathcal{MP}$ in the decision rules (10), we obtain an upper bound on $\mathcal{MP}$, which has the same general structure as problem $\mathcal{P}^u$ in Section 2. Therefore, by using robust optimization techniques, it may be conservatively approximated by the following conic optimization problem:

$$
\begin{aligned}
\widehat{\mathcal{MP}^u}: \quad & \inf \quad \frac{1}{2}\sum_{\rho=1}^{r} \operatorname{tr}\left(\mathbb{E}(\xi\xi^\top \otimes \xi\xi^\top)(Y_\rho \otimes Y_\rho)\right) + \operatorname{tr}\left(\mathbb{E}(\xi\xi^\top)C^\top X\right) \\
& \text{s.t.} \quad X \in \mathcal{X}, Y_1, \ldots, Y_r \in \mathbb{S}^k, Z_\mu \in \mathbb{S}^{k^t} \quad \forall \mu \in \mathcal{M}_t, t \in \mathcal{T} \\
& \qquad \left.\begin{aligned} & P_t^\top \left(\sum_{\kappa,\kappa'=1}^{k} e_\kappa^\top \tilde{A}_\mu X e_{\kappa'} \Sigma_{t\kappa\kappa'} + Z_\mu \right) P_t = \frac{1}{2}\left(e_1 \tilde{b}_\mu^\top + \tilde{b}_\mu e_1^\top\right) \\ & P_t^\top Z_\mu P_t \succeq_{\hat{\mathcal{C}}} 0 \end{aligned}\right\} \quad \forall \mu \in \mathcal{M}_t, \; t \in \mathcal{T} \\
& \qquad \tfrac{1}{2}\left(F_\rho^\top X + X^\top F_\rho\right) - Y_\rho = 0 \quad \forall \rho = 1, \ldots, r \\
& \qquad X \succeq_{\hat{\mathcal{K}}^n} 0.
\end{aligned}
$$

3.2 Dual Problem and Approximation

To obtain a lower bound on $\mathcal{MP}^o$, we start by dualizing problem $\mathcal{MP}$. Using a duality scheme analogous to the one described in Section 2.2, it can be shown that the following quadratic multistage stochastic program is dual to $\mathcal{MP}$:

$$
\begin{aligned}
\mathcal{MD}: \quad & \sup \quad -\mathbb{E}\left(\tfrac{1}{2}v(\xi)^\top v(\xi) + b(\xi)^\top u(\xi)\right) \\
& \text{s.t.} \quad u \in \mathcal{H}, v \in \mathcal{L}_{k,r}, s \in \mathcal{G} \\
& \qquad \left.\begin{aligned} & \mathbb{E}_t\left(a_\nu(\xi)^\top u(\xi) + \tilde{f}_\nu(\xi)^\top v(\xi)\right) - s_\nu(\xi^t) = -c_\nu(\xi^t) \quad \forall \nu \in \mathcal{N}_t, t \in \mathcal{T} \\ & u(\xi) \geq 0, s(\xi) \geq 0 \end{aligned}\right\} \mathbb{P}\text{-a.s.}
\end{aligned}
$$

Next, we require the dual decisions to be representable as

$$
u(\xi) = U\xi, \quad v(\xi) = V\xi \quad \text{and} \quad s_\nu(\xi^t) = (\xi^t)^\top S_\nu \xi^t
$$

for some matrices $V \in \mathbb{R}^{r \times k}$, $S_\nu \in \mathbb{S}^{k^t}, \nu \in \mathcal{N}_t, t \in \mathcal{T}$, and

$$
U = \begin{bmatrix} U_1 P_1 \\ \vdots \\ U_T P_T \end{bmatrix} \tag{11}
$$

for some $U_t \in \mathbb{R}^{m_t \times k^t}, t \in \mathcal{T}$. We denote by $\mathcal{U}$ the linear space of all block triangular matrices of the form (11). With these conventions, problem $\mathcal{MD}$ reduces to a semi-infinite problem similar to $\mathcal{D}^l$, which may be conservatively approximated by the following conic program:

$$
\begin{aligned}
\widehat{\mathcal{MD}^l}: \quad & \sup \quad -\tfrac{1}{2}\operatorname{tr}\left\{V\mathbb{E}\big(\xi\xi^\top\big)V^\top\right\} - \operatorname{tr}\left\{\mathbb{E}\big(\xi\xi^\top\big)B^\top U\right\} \\
& \text{s.t.} \quad U \in \mathcal{U}, V \in \mathbb{R}^{r\times k}, S_\nu \in \mathbb{S}^{k^t} \quad \forall \nu \in \mathcal{N}_t, t \in \mathcal{T} \\
& \qquad \left.\begin{aligned}
& P_t^\top \left(\sum_{\kappa,\kappa'=1}^{k} e_\kappa^\top \big(A_\nu^\top U + \tilde{F}_\nu V\big) e_{\kappa'} \Sigma_{t\kappa\kappa'} - S_\nu \right) P_t = -\frac{1}{2}\Big(e_1 \tilde{c}_\nu^\top + \tilde{c}_\nu e_1^\top\Big) \\
& P_t^\top S_\nu P_t \succeq_{\hat{\mathcal{C}}} 0
\end{aligned}\right\} \forall \nu \in \mathcal{N}_t, \; t \in \mathcal{T} \\
& \qquad U \succeq_{\hat{\mathcal{K}}^m} 0.
\end{aligned}
$$

The following Theorem provides a generalization of Theorem 2.1 to the multistage case.

Theorem 3.1. *In a multistage setting, the following relation holds:*

$$\sup \widehat{\mathcal{MD}^l} \le \sup \mathcal{MD} \le \inf \mathcal{MP} \le \inf \widehat{\mathcal{MP}^u}.$$

The sizes of the conic programs $\widehat{\mathcal{MD}^l}$ and $\widehat{\mathcal{MP}^u}$ are polynomial in k, l, m and n, implying that they are efficiently solvable. The loss of optimality due to the adopted decision rule approximation is bounded by the difference between the optimal values of $\widehat{\mathcal{MD}^l}$ and $\widehat{\mathcal{MP}^u}$.

4 Concluding Remarks

In this paper, we propose tractable primal and dual decision rule approximations for quadratic stochastic programs with random recourse. We have numerically evaluated our approximation scheme in [19, Section 3.4] in the context of a mean-variance portfolio optimization problem. For the problems studied, the best linear-quadratic decision rules are provably optimal to within a few percent only. It is further shown that the popular sample average approximation [3, Section 5.8] fails to solve this problem at a comparable accuracy. The portfolio model further exemplifies the favorable scalability properties of the decision rule approximation as compared to the sample average approximation.

In the future, we plan to evaluate our approximation scheme also in control applications, such as process control and the control of cascaded hydropower systems, that naturally give rise to quadratic stochastic programs. An important direction for future research is the design of more refined approximation schemes to reduce the optimality gap, e.g., by using piecewise linear or higher-order polynomial decision rules. While such decision rules have been investigated in the context of *linear* stochastic programming (see, e.g., [5, 8, 9, 10]), extensions to quadratic stochastic programs have not been studied, and systematic procedures for finding the best approximation (with minimum optimality gap) for a limited budget of computational resources are not yet available.

Acknowledgements: We thank Fundação para a Ciência e a Tecnologia and EPSRC (under grant EP/H0204554/1) for financial support. We are grateful to the referees and the editor for their helpful suggestions for improving the manuscript.

References

1. Van Slyke, R., Wets, R.: L-shaped linear programs with applications to optimal control and stochastic programming. SIAM Journal on Applied Mathematics **17**(4), 638–663 (1969)

2. Römisch, W., Wets, R.: Stability of ε-approximate solutions to convex stochastic programs. SIAM Journal on Optimization **18**(3), 961–979 (2007)

3. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming: Modeling and Theory. SIAM, Philadelphia (2009)

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

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

6. Shapiro, A., Nemirovski, A.: On complexity of stochastic programming problems. Applied Optimization **99**, 111–146 (2005)

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

8. Georghiou, A., Wiesemann, W., Kuhn, D.: Generalized decision rule approximations for stochastic programming via liftings (2010). Available on Optimization Online, submitted for publication

9. Bampou, D., Kuhn, D.: Polynomial approximations for continuous linear programs. SIAM Journal on Optimization **22**(2), 628–648 (2012)

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

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

12. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust Optimization. Princeton University Press, Princeton (2009)

13. Ye, Y.: Interior Point Algorithms: Theory and Analysis. Wiley, New York (1997)

14. Wiesemann, W., Kuhn, D., Rustem, B.: Robust Markov decision processes. Mathematics of Operations Research (Forthcoming)

15. Alizadeh, F., Goldfarb, D.: Second-order cone programming. Mathematical Programming **95**(1), 3–51 (2003)

16. Lobo, M., Vandenberghe, L., Boyd, S., Lebret, H.: Applications of second-order cone programming. Linear Algebra and its Applications **284**(1-3), 193–228 (1998)

17. Wright, S.: Primal-dual aggregation and disaggregation for stochastic linear programs. Mathematics of Operations Research **19**(4), 893–908 (1994)

18. Luenberger, D.: Optimization by Vector Space Methods. Wiley, New York (1969)

19. Rocha, P.: Medium-term planning in deregulated energy markets with decision rules. Ph.D. thesis, Imperial College London (2012)