Barrier Methods Based on Jordan-Hilbert Algebras for Stochastic Optimization in Spin Factors

Baha Alzalg[1,2] *

Abstract We present decomposition logarithmic-barrier interior-point methods based on unital Jordan-Hilbert algebras for infinite-dimensional stochastic second-order cone programming problems in spin factors. The results show that the iteration complexity of the proposed algorithms is independent on the choice of Hilbert spaces from which the underlying spin factors are formed, and so it coincides with the best known complexity obtained by such methods for the finite-dimensional setting. We apply our results to an important problem in stochastic control, namely the two-stage stochastic multi-criteria design problem. We show that the corresponding infinite-dimensional system in this case is a matrix differential Ricatti equation plus a finite-dimensional system, and hence, it can be solved efficiently to find the search direction.

Keywords Jordan-Hilbert algebras · Second-order cone programming · Programming in abstract spaces · Stochastic programming · Interior-point methods · Stochastic control

Mathematics Subject Classification 17C65 · 90C15 · 90C25 · 90C48 · 90C51 · 93E20

1 Introduction

Many authors studied optimization problems in general spaces (see for example [1–10]). In several practical problems, such as stochastic control, optimization of stochastic systems in continuous time, control theory, etc., we find diverse application models lead to stochastic convex optimization problems in general spaces. This motivates us to study stochastic programs over infinite-dimensional convex cones. As an important paradigm of such programs, in this paper, we introduce and study two-stage stochastic conic optimization problems in the so-called spin factors. Stochastic conic programming problems in spin factors are stochastic convex optimization problems in which we minimize a linear objective function over the intersection of an affine linear manifold with a direct product of infinite-dimensional second-order cones in spin factors.

Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a real Hilbert space with an inner product $\langle\cdot,\cdot\rangle$. The Hilbert space direct product $\mathcal{S}_{\mathcal{H}} := \mathbb{R} \otimes \mathcal{H}$ is called a *spin factor*. For each element $\boldsymbol{x} \in \mathcal{S}_{\mathcal{H}}$, we write x_0 for the component that belongs to $\mathbb{R}$ and $\bar{\boldsymbol{x}}$ for the subelement that belongs to $\mathcal{H}$. Therefore $\boldsymbol{x} = (x_0, \bar{\boldsymbol{x}}) \in \mathcal{S}_{\mathcal{H}}$. We equip $\mathcal{S}_{\mathcal{H}}$ with the inner product "$\bullet$" which maps the direct product $\mathcal{S}_{\mathcal{H}} \otimes \mathcal{S}_{\mathcal{H}}$ into $\mathbb{R}$ by $\boldsymbol{x} \bullet \boldsymbol{y} := x_0 y_0 + \langle \bar{\boldsymbol{x}}, \bar{\boldsymbol{y}} \rangle$, for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}_{\mathcal{H}}$.

The *infinite-dimensional second-order cone in the spin factor* $\mathcal{S}_{\mathcal{H}}$ is defined as

$$\mathcal{K}_{\mathcal{H}} := \left\{ \boldsymbol{x} \in \mathcal{S}_{\mathcal{H}} : \ x_0 \geq \|\bar{\boldsymbol{x}}\| \right\}, \tag{1}$$

where $\|\cdot\| := \sqrt{\langle\cdot,\cdot\rangle}$ is the norm in the Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$. Clearly, when $\mathcal{H} = \mathbb{R}^{n-1}$, for an integer $n > 1$, the norm $\|\cdot\|$ reduces to the Euclidean norm, and the cone $\mathcal{K}_{\mathcal{H}}$ reduces to the well-known *n^{th}-dimensional second-order cone in the Euclidean space* $\mathbb{R} \times \mathbb{R}^{n-1}$, which is defined as $\mathcal{Q} := \{\boldsymbol{x} \in \mathbb{R} \times \mathbb{R}^{n-1} : \ x_0 \geq \|\bar{\boldsymbol{x}}\|\} = \mathcal{K}_{\mathbb{R}^{n-1}}$. Hence, the stochastic conic optimization problem we study in this paper includes the finite-dimensional stochastic second-order cone programming problem as a special case.

✉ * B. Alzalg (b.alzalg@ju.edu.jo)

1 Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43210, USA
2 Department of Mathematics, The University of Jordan, Amman 11942, Jordan

Logarithmic-barrier Benders decomposition interior-point algorithms based on Euclidean Jordan algebras have been derived in [11] for solving finite-dimensional stochastic second-order cone programming with recourse. Inspired by the work in [12], which asserts the one-to-one correspondence between the unital Jordan-Hilbert algebras (JH-algebras, for short) and the infinite-dimensional symmetric cones, it is our purpose in this paper to extend the above logarithmic-barrier Benders decomposition methods and their self-concordance analysis to the infinite-dimensional setting by deriving logarithmic-barrier Benders decomposition interior-point methods based on unital JH-algebras for *two-stage stochastic conic programming problem in spin factors with K scenarios*:

$$\begin{array}{lll} \min & \boldsymbol{c} \bullet \boldsymbol{x} + \sum_{k=1}^{K} \bar{\rho}^{(k)}(\boldsymbol{x}) & \text{where } \bar{\rho}^{(k)}(\boldsymbol{x}) \text{ is} \\ \text{s.t.} & \boldsymbol{a}_i \bullet \boldsymbol{x} = b_i,\ i = 1, \ldots, m_1, & \text{the minimum} \\ & \boldsymbol{x} \in \mathcal{G} \bigcap \bigotimes_{i=1}^{r_1} \mathcal{K}_{\mathcal{H}_i}, & \text{value of the problem} \end{array} \quad \begin{array}{ll} \min & \boldsymbol{d}^{(k)} \bullet \boldsymbol{y}^{(k)} \\ \text{s.t.} & \boldsymbol{w}_j^{(k)} \bullet \boldsymbol{y}^{(k)} = h_j^{(k)} - \boldsymbol{t}_j^{(k)} \bullet \boldsymbol{x},\ j = 1, \ldots, m_2, \\ & \boldsymbol{y}^{(k)} \in \hat{\mathcal{G}} \bigcap \bigotimes_{j=1}^{r_2} \mathcal{K}_{\hat{\mathcal{H}}_j}, \end{array} \tag{2}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}^{(k)}$ are the first- and second-stage decision variables, $\mathcal{G}$ and $\hat{\mathcal{G}}$ are closed subspaces in direct products of given spin factors, $\bigotimes_{i=1}^{r_1} \mathcal{K}_{\mathcal{H}_i}$ and $\bigotimes_{j=1}^{r_2} \mathcal{K}_{\hat{\mathcal{H}}_j}$ are direct products of infinite-dimensional second-order cones in the corresponding spin factors, and "$\bullet$" is an inner product that is induced by the underlying spin factors and will be appropriate for the set of data given above. For simplicity, we used the same symbol to denote the inner product in the first-stage problem in (2) and that in the second-stage problem in (2). We also assumed that $\boldsymbol{d}^{(k)}$ has already absorbed the scenario probabilities for each $k = 1, 2, \ldots, K$. The function $\bar{\rho}^{(k)}(\boldsymbol{x})$ is called the *recourse function*.

In recent developments, besides the algorithm developed in [11], there are other interior-point algorithms proposed for solving the finite-dimensional stochastic second-order cone programming. We briefly outline these results ordered chronologically. In 2014, a homogeneous self-dual interior-point algorithm was proposed in [13] for solving the problem. Later in the same year, a decomposition-based interior-point algorithm based on a *logarithmic* barrier was proposed in [11] for solving the problem (see also [14]). In 2015, a decomposition-based interior-point algorithm based on a *volumetric* barrier was proposed in [15] for solving the problem. In 2018, an infeasible self-dual interior-point algorithm was proposed in [16] for solving the problem. We emphasize that all interior-point algorithms in [11,13,15,16] have been developed and analyzed for the finite-dimensional stochastic second-order cone programming. Because there are no infinite-dimensional analogs of the interior-point methods proposed in [11,13,15,16], a novel and workable extension to such methods is in turning from finite-dimensional Euclidean space $\mathbb{R} \times \mathbb{R}^{n-1}$ to the spin factor space $\mathbb{R} \otimes \mathcal{H}$, which is known to be the most natural generalization of the underlying optimization problem to the infinite-dimensional setting.

After generalizing the logarithmic-barrier Benders decomposition interior-point algorithms to the spin factor space, it will become clear how to generalize the other interior-point algorithms in [13,15,16] analyzed earlier in the finite-dimensional Euclidean space. As a curiosity, the reader may be interested in knowing the reason why we choose to extend specifically logarithmic-barrier Benders decomposition methods. In fact, there are two basic formulations in stochastic programming. The first one is the deterministic equivalence, which is the extensive formulation of a stochastic program that forms an equivalent large one-stage problem containing all constraints and all scenarios. The methods proposed in [13,16] are based on this formulation. The second formulation is based on Benders decomposition, which decomposes a stochastic program into stages where, at each stage, variables at preceding stages are considered as constraints so that the subproblem at the current stage is easier to solve. The methods proposed in [11, 15] are based in this formulation. Such methods do not require explicit knowledge of all the scenarios in the algorithm up front because the scenarios can be added as the algorithm progresses. This has a great advantage over the other methods in [13,16] for speeding up the algorithm in its early stages. Adding this to the fact that working on the logarithmic barrier is a necessary step that should precede working on the volumetric barrier, it shows that our selection of the algorithm proposed in [11] is not random, but based on proper criteria.

The crucial difference between finite and infinite-dimensional settings is in solving the optimality conditions to find the Newton-type search directions, which need to be found in each iteration. After the seminal work of Kantorovich (1942), it became clear that Newton's method can be applied in the infinite-dimensional case, but the procedure of the Newton-type search directions requires a separate analysis for each infinite-dimensional problem formulation. From this stems the importance of the results obtained in this paper. Furthermore, the existence of the Fréchet derivative (which is the derivative on Banach spaces) and the existence of the central path (which is the path that minimizes the barrier function) are nontrivial issues in an infinite-dimensional situation. Since it is easier to contemplate this with an example, rather than a general setting, in this paper we apply our results to an important example in stochastic control: the two-stage stochastic multi-criteria design problem. After presenting the proposed algorithms, we also consider short- and long-step versions of the algorithms and obtain their iteration complexity results. We will see that these complexity results coincide with the best known in the finite-dimensional case. Most importantly, the complexity of the proposed algorithms can seen to be independent on the choice of Hilbert spaces from which the underlying spin factors are formed. Herein also lies one of the novelty of this paper.

The generalization to the general spin factor case is not trivial, however, but challenging. The reason for that is not only because the notion of the Fréchet derivative is highly involved in our computation instead of the usual derivative, but also because the infinite-dimensional Jordan algebra of spin factors is not fully established in the literature like the finite-dimensional Euclidean Jordan algebra. In addition, while this paper is also influenced, in part, by the pioneer work of Faybusovich and Tsuchiya [8] for infinite-dimensional *deterministic* second-order cone programming, the stochasticity adds a dimension of difficulty and a layer of complexity not previously associated with this generalization.

The main difference between the stochastic optimization problem we consider in this paper and that considered in [17] is that, while both in a Hilbert space, the optimization problem we consider here is introduced in spin factors, in which all constraints are infinite-dimensional second-order conic constraints, while the one considered in [17] is a stochastic linear optimization problem, in which all constraints are linear, but its variables are launched to be elements in an infinite-dimensional Hilbert space instead of being vectors in $\mathbb{R}^n$. It is our firm belief that each of these two cases (the case in this paper and that in [17]) deserves its own study.

This paper is organized by section as follows. In Section 2, we give preliminaries of spin factor Jordan algebras and present the JH-algebraic structure of the infinite-dimensional second-order cones. In Section 3, we introduce the problem formulation, some assumptions and the associated barrier problem. Section 4 is devoted to compute the Fréchet derivatives of the recourse function. In Section 5, we present some fundamental properties of the recourse function. More specifically, we show that the set of barrier functions for positive values of barrier parameter comprises a self-concordant family. Based on such fundamental properties, the proposed interior-point decomposition algorithm is presented in Section 6. The complexity results for short- and long-step versions of the proposed algorithm are presented with proofs in Section 7. In Section 8, we apply our results to the two-stage stochastic multi-criteria design problem. Section 9 contains concluding remarks. Appendix A is a complement to Section 2. It reviews basics of the unital JH-algebras, and presents the properties associated with the algebra of the underlying cone. In Appendix B, we state some technical lemmas which are needed in proofs of the complexity analysis of the proposed algorithms.

2 Jordan-Hilbert algebraic structure

In this section, we present the JH-algebraic structure of the cone $\mathcal{K}$ defined in (1). The reader is also invited to read the material presented in Appendix A, which complements and extends this section.

Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a real Hilbert space, which can be infinite-dimensional, with an inner product $\langle\cdot,\cdot\rangle$, identity operator $\mathcal{I}_{\mathcal{H}}$, and zero element $\mathbf{0}$. Then $\mathcal{S} := \mathcal{S}_{\mathcal{H}} = \mathbb{R} \otimes \mathcal{H}$ is the corresponding spin factor. The *orthogonal complement* of a closed subspace $\mathcal{G}$ of $\mathcal{S}$ is the subspace $\mathcal{G}^\perp := \{\boldsymbol{x} \in \mathcal{S} : \boldsymbol{x} \perp \mathcal{G}\}$.

For $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$, we define the maps $\bullet : \mathcal{S} \otimes \mathcal{S} \longrightarrow \mathbb{R}$ and $\circ : \mathcal{S} \otimes \mathcal{S} \longrightarrow \mathcal{S}$ as

$$\boldsymbol{x} \bullet \boldsymbol{y} := x_0 y_0 + \langle \bar{x}, \bar{y} \rangle, \ \text{ and } \ \boldsymbol{x} \circ \boldsymbol{y} := (\boldsymbol{x} \bullet \boldsymbol{y}, x_0 \bar{y} + y_0 \bar{x}), \tag{3}$$

respectively. The structure $(\mathcal{S}, \circ)$ is an JH-algebra (see Definition 3), and the element $\boldsymbol{e} := (1, \mathbf{0})$ is called the *identity* of $(\mathcal{S}, \circ)$. The fundamental result in Theorem 7 states that the cone of squares of $\mathcal{S}$, which is defined as $\mathcal{S}^+ := \{\boldsymbol{x} \circ \boldsymbol{x} : \boldsymbol{x} \in \mathcal{S}\}$, is the infinite-dimensional second-order cone $\mathcal{K}$ defined in (1).

The *spectral decomposition* associated with each element $\boldsymbol{x} \in \mathcal{S}$ is the decomposition

$$\boldsymbol{x} = \underbrace{(x_0 + \|\bar{x}\|)}_{\lambda_1(\boldsymbol{x})} \underbrace{\left(\frac{1}{2}\right)\left(1, \frac{\bar{x}}{\|\bar{x}\|}\right)}_{\boldsymbol{c}_1(\boldsymbol{x})} + \underbrace{(x_0 - \|\bar{x}\|)}_{\lambda_2(\boldsymbol{x})} \underbrace{\left(\frac{1}{2}\right)\left(1, \frac{-\bar{x}}{\|\bar{x}\|}\right)}_{\boldsymbol{c}_2(\boldsymbol{x})}.$$

The values $\lambda_1(x)$ and $\lambda_2(x)$ are called the *eigenvalues* of x, and the vectors $c_1(x)$ and $c_2(x)$ are called the *eigenelements* of x. Since each $x \in \mathcal{S}$ has 2 eigenvalues (including multiplicities), we say that the JH-algebra $\mathcal{S}$ is of *rank* two. Given the above spectral decomposition, the *trace* and *determinant* are defined in $(\mathcal{S}, \circ)$ as $\text{trace}(x) := \lambda_1(x) + \lambda_2(x) = 2x_0$, and $\det(x) := \lambda_1(x)\lambda_2(x) = x_0^2 - \|\bar{x}\|^2$, respectively.

We use $\mathcal{I}$ and $\mathcal{R}$ to denote the *identity* and *reflection operators* in $\mathcal{S}$ respectively, which are defined by the block partitions

$$\mathcal{I} := \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_{\mathcal{H}} \end{pmatrix}, \text{ and } \mathcal{R} := \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & -\mathcal{I}_{\mathcal{H}} \end{pmatrix}.$$

For a smooth function $f(\cdot)$, we define the image of $x \in \mathcal{S}$ under f as $f(x) := f(\lambda_1(x))\, c_1(x) + f(\lambda_2(x))\, c_2(x)$. In particular, the *inverse* of $x \in \text{int}\, \mathcal{K}$ (for which $\det(x) \neq 0$) is given by

$$x^{-1} := \frac{1}{\lambda_1(x)} c_1(x) + \frac{1}{\lambda_2(x)} c_2(x) = \frac{1}{\det(x)} (x_0, -\bar{x}) = \frac{1}{\det(x)} \mathcal{R} x.$$

One can show that $x \circ x^{-1} = e$.

We now describe the *linear multiplication operator* $\mathcal{L} : \mathcal{S} \longrightarrow \mathcal{S}$. Associated with each element $x = (x_0, \bar{x}) \in \mathcal{S}$, we introduce the continuous linear maps

$$\begin{array}{llll} \bar{\mathcal{X}} : \mathcal{H} \longrightarrow \mathbb{R} & \text{defined by} & \bar{\mathcal{X}}\zeta := \langle \bar{x}, \zeta \rangle, & \text{for} \quad \zeta \in \mathcal{H}, \\ \bar{\mathcal{X}}^{\dagger} : \mathbb{R} \longrightarrow \mathcal{H} & \text{defined by} & \bar{\mathcal{X}}^{\dagger}\alpha := \alpha \bar{x}, & \text{for} \quad \alpha \in \mathbb{R}. \end{array}$$

Accordingly, associated with elements $x = (x_0, \bar{x}) \in \mathcal{S}$ and $z = (z_0, \bar{z}) \in \mathcal{S}$, we also introduce the continuous quadratic maps

$$\begin{array}{llll} \bar{\mathcal{X}}^{\dagger}\bar{\mathcal{Z}} : \mathcal{H} \longrightarrow \mathcal{H} & \text{defined by} & \bar{\mathcal{X}}^{\dagger}\bar{\mathcal{Z}}\zeta := \langle \bar{z}, \zeta \rangle \bar{x}, & \text{for} \quad \zeta \in \mathcal{H}, \\ \mathcal{X}^{\dagger}\mathcal{Z} : \mathcal{S} \longrightarrow \mathcal{S} & \text{defined by the block partition} & \mathcal{X}^{\dagger}\mathcal{Z} := \begin{pmatrix} x_0 z_0 & x_0 \bar{\mathcal{Z}} \\ z_0 \bar{\mathcal{X}}^{\dagger} & \bar{\mathcal{X}}^{\dagger}\bar{\mathcal{Z}} \end{pmatrix}. & \end{array}$$

Thinking of each element $x := (x_0, \bar{x}) \in \mathcal{S}$ as a "column vector" $\begin{pmatrix} x_0 \\ \bar{x} \end{pmatrix}$, the linear operator $\mathcal{L}(\cdot)$ for the JH-algebra $(\mathcal{S}, \circ)$ admits the following block partition

$$\mathcal{L}(x) := \begin{pmatrix} x_0 & \bar{\mathcal{X}} \\ \bar{\mathcal{X}}^{\dagger} & x_0 \mathcal{I}_{\mathcal{H}} \end{pmatrix}.$$

This gives the explicit formula of $\mathcal{L}(\cdot)$ as introduced in [8, Proposition 10]. We immediately have

$$\mathcal{L}(x)y = \begin{pmatrix} x_0 & \bar{\mathcal{X}} \\ \bar{\mathcal{X}}^{\dagger} & x_0 \mathcal{I}_{\mathcal{H}} \end{pmatrix} \begin{pmatrix} y_0 \\ \bar{y} \end{pmatrix} = \begin{pmatrix} x_0 y_0 + \bar{\mathcal{X}}\bar{y} \\ \bar{\mathcal{X}}^{\dagger} y_0 + x_0 \mathcal{I}_{\mathcal{H}} \bar{y} \end{pmatrix} = \begin{pmatrix} x_0 y_0 + \langle \bar{y}, \bar{x} \rangle \\ y_0 \bar{x} + x_0 \bar{y} \end{pmatrix} = x \circ y.$$

In particular, $\mathcal{L}(x)e = x$ and $\mathcal{L}(x)x = x^2$.

The *quadratic representation* $\mathcal{Q}_{\cdot} : \mathcal{S} \longrightarrow \mathcal{S}$ of $x \in (\mathcal{S}, \circ)$ is given by (cf. [8, Proposition 14])

$$\mathcal{Q}_x := 2\mathcal{L}^2(x) - \mathcal{L}(x^2) = \begin{pmatrix} x \bullet x & 2x_0 \bar{\mathcal{X}} \\ 2x_0 \bar{\mathcal{X}}^{\dagger} & \det(x)\mathcal{I}_{\mathcal{H}} + 2\bar{\mathcal{X}}^{\dagger}\bar{\mathcal{X}} \end{pmatrix} = 2\mathcal{X}^{\dagger}\mathcal{X} - \det(x)\mathcal{R}.$$

Thus $\mathcal{Q}_x y = 2\langle x, y \rangle x - \det(x)\mathcal{R}y$ for $y \in \mathcal{S}$. One can see that $\mathcal{Q}_e = \mathcal{I}$, $\mathcal{Q}_x e = x^2$, $\mathcal{Q}_x x^{-1} = x$, and $\mathcal{Q}_{x^{-1}} = \mathcal{Q}_x^{-1}$.

The *quadratic operator* for the algebra $(\mathcal{S}, \circ)$ is accordingly given by

$$\mathcal{Q}_{x,z} := \mathcal{L}(x)\mathcal{L}(z) + \mathcal{L}(z)\mathcal{L}(x) - \mathcal{L}(x \circ z) = \begin{pmatrix} x \bullet z & x_0 \bar{\mathcal{Z}} + z_0 \bar{\mathcal{X}} \\ x_0 \bar{\mathcal{Z}}^{\dagger} + z_0 \bar{\mathcal{X}}^{\dagger} & (x_0 z_0 - \langle \bar{x}, \bar{z} \rangle)\, \mathcal{I}_{\mathcal{H}} + \bar{\mathcal{X}}^{\dagger}\bar{\mathcal{Z}} + \bar{\mathcal{Z}}^{\dagger}\bar{\mathcal{X}} \end{pmatrix}.$$

Let $x \in (\mathcal{S}, \circ)$. The *Frobenius norm* and *2-norm* of $x \in \mathcal{S}$ are defined as

$$\|x\|_F := \sqrt{\lambda_1^2(x) + \lambda_2^2(x)} = \sqrt{2\, x \bullet x} \text{ and } \|x\|_2 := \max\{|\lambda_1(x)|, |\lambda_2(x)|\} = |x_0| + \|\bar{x}\|,$$

respectively. It is clear that $\|x\|_2 \leq \|x\|_F$. One can also prove that

$$|x \bullet y| \leq \|x\|_F \|y\|_F \quad \text{and} \quad \|x \circ y\|_F \leq \|x\|_F \|y\|_F.$$

An operator $\mathcal{T} : \mathcal{S} \longrightarrow \mathcal{S}$ is said to be *invertible* if there exists another operator, called the *inverse operator* of $\mathcal{T}$ and denoted by $\mathcal{T}^{-1} : \mathcal{S} \longrightarrow \mathcal{S}$, such that $\mathcal{T}\mathcal{T}^{-1} = \mathcal{T}^{-1}\mathcal{T} = \mathcal{I}$.

The above notions and concepts are also used in the block sense as follows. Let $(\mathcal{S}_1, \circ_1, \bullet_1)$, $(\mathcal{S}_2, \circ_2, \bullet_2), \ldots,$ $(\mathcal{S}_r, \circ_r, \bullet_r)$ be spin factors with identity elements $e_1, e_2, \ldots, e_r$, identity operators $\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_r$, reflection operators $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_r$, and cones of squares $\mathcal{S}_1^+, \mathcal{S}_2^+, \ldots, \mathcal{S}_r^+$, respectively. Let also $x := (x_1, x_2, \ldots, x_r)$ and $y := (y_1, y_2, \ldots, y_r)$, with $x_i, y_i \in (\mathcal{S}_i, \circ_i, \bullet_i)$ for $i = 1, 2, \ldots, r$. Then[1]

$\mathcal{S}$	:=	$\mathcal{S}_1 \otimes \mathcal{S}_2 \otimes \cdots \otimes \mathcal{S}_r$ is a direct product of spin factors;
$\mathcal{S}^+$	:=	$\mathcal{S}_1^+ \otimes \mathcal{S}_2^+ \otimes \cdots \otimes \mathcal{S}_r^+$ is the cone of squares of $\mathcal{S}$;
e	:=	$(e_1, e_2, \ldots, e_r)$ is the identity element of $\mathcal{S}$;
$x \circ y$	:=	$(x_1 \circ_1 y_1, x_2 \circ_2 y_2, \ldots, x_r \circ_r y_r)$ is the Jordan product in $\mathcal{S}$;
$x \bullet y$	:=	$x_1 \bullet_1 y_1 + x_2 \bullet_2 y_2 + \cdots + x_r \bullet_r y_r$ is the inner product in $\mathcal{S}$;
x^{-1}	:=	$\left(x_1{}^{-1}, x_2{}^{-1}, \ldots, x_r{}^{-1}\right)$ is the inverse of x (provided x_i is invertible for each i);
$f(x)$	:=	$(f(x_1), f(x_2), \ldots, f(x_r))$. (Here f is a smooth function);
$\text{trace}(x)$	:=	$\text{trace}(x_1) + \text{trace}(x_2) + \cdots + \text{trace}(x_r)$;
$\det(x)$	:=	$\det(x_1)\det(x_2)\cdots\det(x_r)$;
$\mathcal{I}$	:=	$\mathcal{I}_1 \oplus \mathcal{I}_2 \oplus \cdots \oplus \mathcal{I}_r$ is the identity operator in $\mathcal{S}$;
$\mathcal{R}$	:=	$\mathcal{R}_1 \oplus \mathcal{R}_2 \oplus \cdots \oplus \mathcal{R}_r$ is the reflection operator in $\mathcal{S}$;
$\mathcal{L}(x)$	:=	$\mathcal{L}(x_1) \oplus \mathcal{L}(x_2) \oplus \cdots \oplus \mathcal{L}(x_r)$;
$\mathcal{Q}_x$	:=	$\mathcal{Q}_{x_1} \oplus \mathcal{Q}_{x_2} \oplus \cdots \oplus \mathcal{Q}_{x_r}$;
$\mathcal{Q}_{x,y}$	:=	$\mathcal{Q}_{x_1,y_1} \oplus \mathcal{Q}_{x_2,y_2} \oplus \cdots \oplus \mathcal{Q}_{x_r,y_r}$;
$\|x\|_F^2$	:=	$\|x_1\|_F^2 + \|x_2\|_F^2 + \cdots + \|x_r\|_F^2$;
$\|x\|_2$	:=	$\max\{\|x_1\|_2, \|x_2\|_2, \ldots, \|x_r\|_2\}$.

Note that $x \in \mathcal{S}$ has $2r$ eigenvalues (including multiplicities) comprised of the union of the eigenvalues of each $x_i \in \mathcal{S}_i$ for $i = 1, 2, \ldots, r$. The value $2r$ is called the *rank* of $\mathcal{S}$.

We write the multiple-block conic inequality $x \succeq \mathbf{0}$ ($x \succ \mathbf{0}$) to mean that $x \in \mathcal{S}^+$ ($x \in \text{int}\,\mathcal{S}^+$). It is immediately seen that, for every element $x \in \mathcal{S}$, $x \succeq \mathbf{0}$ if and only if x is partitioned conformally as $x = (x_1; x_2; \ldots; x_r)$ with $x_i \succeq \mathbf{0}$ for $i = 1, 2, \ldots, r$. We write $x \succeq y$ or $y \preceq x$ to mean that $x - y \succeq \mathbf{0}$. We also write $X \succeq 0$ ($X \succ 0$) to mean that X is a symmetric positive semidefinite (positive definite) matrix.

Throughout the paper we drop the subscript $\mathcal{H}$ from $\mathcal{S}_{\mathcal{H}}$ and $\mathcal{K}_{\mathcal{H}}$ and simply write $\mathcal{S}$ and $\mathcal{K}$, respectively, when they are clear from the context. We also use "$\mathcal{D}_x$", "$\mathcal{D}_{xx}^2$" and "$\mathcal{D}_{xxx}^3$" to denote the first, second and third Fréchet derivatives with respect to x, respectively. The above derivatives are defined in the standard way of this sense (see also [10, Section 2]): If $f : \mathcal{S} \longrightarrow \mathbb{R}$ is a smooth function, the gradient $\nabla_x f(x)$ is uniquely determined by

$$\mathcal{D}_x f(x)(\xi) := \nabla_x f(x) \bullet \xi$$

for any $\xi \in \mathcal{S}$. Here, $\mathcal{D}_x f(x)(\xi)$ stands for the first Fréchet derivative of f at the element x evaluated on ξ. The second Fréchet derivative is the derivative of the first Fréchet derivative, which is given by

$$\mathcal{D}_{xx}^2 f(x)(\xi, \zeta) := \mathcal{D}_x\left(\mathcal{D}_x f(x)(\xi)\right)(\zeta)$$

for any $\xi, \zeta \in \mathcal{S}$. Here, $\mathcal{D}_{xx}^2 f(x)(\xi, \zeta)$ stands for the second Fréchet derivative of f at the element x evaluated on (ξ, ζ). Higher Fréchet derivatives are defined in a similar way.

3 Formulation

In this section, we introduce the problem formulation and the two-stage barrier problem, give some assumptions, and present the commutative directions.

[1] The direct sum of two block partitions $\mathcal{A}$ and $\mathcal{B}$ is the block diagonal partition $\mathcal{A} \oplus \mathcal{B} := \begin{pmatrix} \mathcal{A} & 0 \\ 0 & \mathcal{B} \end{pmatrix}$.

Based on the notations given in Section 2, we write $\mathcal{K}$ in Problem (2) as $\mathcal{K} = \mathcal{S}^+$ where $(\mathcal{S}, \circ, \bullet)$ is a direct product of r_1 spin factors with identity element $\boldsymbol{e}$, identity operator $\mathcal{I}$, reflection operator $\mathcal{R}$, closed subspace $\mathcal{G}$, and cone of squares $\mathcal{S}^+$. We also write $\hat{\mathcal{K}}$ in Problem (2) as $\hat{\mathcal{K}} = \hat{\mathcal{S}}^+$ where $(\hat{\mathcal{S}}, \circ, \bullet)$ is a direct product of r_2 spin factors with identity element $\hat{\boldsymbol{e}}$, identity operator $\hat{\mathcal{I}}$, reflection operator $\hat{\mathcal{R}}$, closed subspace $\hat{\mathcal{G}}$, and cone of squares $\hat{\mathcal{S}}^+$. Here, for the sake of simplicity, we denoted inner products and Jordan products on $\mathcal{S}$ and $\hat{\mathcal{S}}$ by the same symbols. Also, for simplicity, we will denote conic inequalities on $\mathcal{S}$ and $\hat{\mathcal{S}}$ by the same symbols "$\succeq$" and "$\succ$" because the context will show their constrained cone distinctly.

From the context of Problem (2), it is pretty clear that $\boldsymbol{c} \in \mathcal{S}, \boldsymbol{a}_i \in \mathcal{S}, \boldsymbol{t}_j^{(k)} \in \mathcal{S}, \boldsymbol{x} \in \mathcal{S}, \boldsymbol{b} \in \mathbb{R}^{m_1}$, $\boldsymbol{d}^{(k)} \in \hat{\mathcal{S}}$, $\boldsymbol{w}_j^{(k)} \in \hat{\mathcal{S}}, \boldsymbol{y}^{(k)} \in \hat{\mathcal{S}}$ and $\boldsymbol{h}^{(k)} \in \mathbb{R}^{m_2}$ for $i = 1, 2, \ldots, m_1$, $j = 1, 2, \ldots, m_2$ and $k = 1, 2, \ldots, K$. Throughout the rest of the paper, we define the maps $\mathcal{A} : \mathcal{S} \longrightarrow \mathbb{R}^{m_1}, \mathcal{T}^{(k)} : \mathcal{S} \longrightarrow \mathbb{R}^{m_2}$ and $\mathcal{W}^{(k)} : \hat{\mathcal{S}} \longrightarrow \mathbb{R}^{m_2}$ by

$$\mathcal{A}\boldsymbol{x} := \begin{bmatrix} \boldsymbol{a}_1 \bullet \boldsymbol{x} \\ \boldsymbol{a}_2 \bullet \boldsymbol{x} \\ \vdots \\ \boldsymbol{a}_{m_1} \bullet \boldsymbol{x} \end{bmatrix}, \ \mathcal{T}^{(k)}\boldsymbol{x} := \begin{bmatrix} \boldsymbol{t}_1^{(k)} \bullet \boldsymbol{x} \\ \boldsymbol{t}_2^{(k)} \bullet \boldsymbol{x} \\ \vdots \\ \boldsymbol{t}_{m_2}^{(k)} \bullet \boldsymbol{x} \end{bmatrix}, \ \text{and} \ \mathcal{W}^{(k)}\boldsymbol{y}^{(k)} := \begin{bmatrix} \boldsymbol{w}_1^{(k)} \bullet \boldsymbol{y}^{(k)} \\ \boldsymbol{w}_2^{(k)} \bullet \boldsymbol{y}^{(k)} \\ \vdots \\ \boldsymbol{w}_{m_2}^{(k)} \bullet \boldsymbol{y}^{(k)} \end{bmatrix}, \ \text{for } \boldsymbol{x} \in \mathcal{S} \text{ and } \boldsymbol{y}^{(k)} \in \hat{\mathcal{S}}, 1 \le k \le K.$$

We also define the maps

$$\begin{array}{llllll} \mathcal{A}^\dagger : \mathbb{R}^{m_1} \longrightarrow \mathcal{S} & \text{by} & \mathcal{A}^\dagger \boldsymbol{\lambda} & := & \sum_{i=1}^{m_1} \lambda_i \boldsymbol{a}_i, & \text{for} \ \ \boldsymbol{\lambda} \in \mathbb{R}^{m_1}, \\ \mathcal{T}^{(k)\dagger} : \mathbb{R}^{m_2} \longrightarrow \mathcal{S} & \text{by} & \mathcal{T}^{(k)\dagger}\boldsymbol{z}^{(k)} & := & \sum_{j=1}^{m_2} z_j^{(k)} \boldsymbol{t}_j^{(k)}, & \text{for} \ \ \boldsymbol{z}^{(k)} \in \mathbb{R}^{m_2}, \ \ k = 1, 2, \ldots, K, \\ \mathcal{W}^{(k)\dagger} : \mathbb{R}^{m_2} \longrightarrow \hat{\mathcal{S}} & \text{by} & \mathcal{W}^{(k)\dagger}\boldsymbol{z}^{(k)} & := & \sum_{j=1}^{m_2} z_j^{(k)} \boldsymbol{w}_j^{(k)}, & \text{for} \ \ \boldsymbol{z}^{(k)} \in \mathbb{R}^{m_2}, \ \ k = 1, 2, \ldots, K. \end{array}$$

Then, we can rewrite Problem (2) as

$$\begin{array}{ll} \min & \boldsymbol{c} \bullet \boldsymbol{x} + \sum_{k=1}^K \bar{\rho}^{(k)}(\boldsymbol{x}) \\ \text{s.t.} & \mathcal{A}\boldsymbol{x} = \boldsymbol{b}, \\ & \boldsymbol{x} \in \mathcal{G}, \boldsymbol{x} \succeq \boldsymbol{0}, \end{array} \quad \begin{array}{l} \text{where } \bar{\rho}^{(k)}(\boldsymbol{x}) \text{ is} \\ \text{the minimum} \\ \text{of the problem} \end{array} \quad \begin{array}{ll} \min & \boldsymbol{d}^{(k)} \bullet \boldsymbol{y}^{(k)} \\ \text{s.t.} & \mathcal{W}^{(k)}\boldsymbol{y}^{(k)} = \boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x}, \\ & \boldsymbol{y}^{(k)} \in \hat{\mathcal{G}}, \boldsymbol{y}^{(k)} \succeq \boldsymbol{0}. \end{array} \tag{4}$$

For $k = 1, 2, \ldots, K$, the dual of the second-stage problem in (4) is

$$\begin{array}{ll} \max & \left(\boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x}\right)^\mathsf{T} \boldsymbol{z}^{(k)} \\ \text{s.t.} & \mathcal{W}^{(k)\dagger}\boldsymbol{z}^{(k)} + \boldsymbol{s}^{(k)} = \boldsymbol{d}^{(k)}, \\ & \boldsymbol{s}^{(k)} \in \hat{\mathcal{G}}^\perp, \boldsymbol{s}^{(k)} \succeq \boldsymbol{0}, \end{array} \tag{5}$$

where $\boldsymbol{z}^{(k)} \in \mathbb{R}^{m_2}$ is the first-stage dual multiplier and $\boldsymbol{s}^{(k)} \in \hat{\mathcal{S}}$ is the second-stage dual slack variable.

Observe that the optimality conditions for $\boldsymbol{y}^{(k)}$ and $(\boldsymbol{z}^{(k)}, \boldsymbol{s}^{(k)})$ to be optimal for the second-stage problem in (4) and Problem (5), respectively, are

$$\begin{array}{l} \boldsymbol{y}^{(k)} \circ \boldsymbol{s}^{(k)} = \boldsymbol{0} \ \left(\text{or equivalently, } \boldsymbol{y}^{(k)} \in \hat{\mathcal{G}} \ \& \ \boldsymbol{s}^{(k)} \in \hat{\mathcal{G}}^\perp\right), \\ \mathcal{W}^{(k)}\boldsymbol{y}^{(k)} = \boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x}, \\ \mathcal{W}^{(k)\dagger}\boldsymbol{z}^{(k)} + \boldsymbol{s}^{(k)} = \boldsymbol{d}^{(k)}, \\ \boldsymbol{y}^{(k)} \succeq \boldsymbol{0}, \ \boldsymbol{s}^{(k)} \succeq \boldsymbol{0}. \end{array}$$

We assume the fulfillment of constraint qualifications to ensure the validity of the above optimality conditions (see Assumptions 1 and 2). Let $\mu > 0$ be a barrier parameter. The *two-stage log-barrier problem* associated with Problem (4) is

$$\begin{array}{ll} \min & \boldsymbol{c} \bullet \boldsymbol{x} - \mu \ln\det \boldsymbol{x} + \sum_{k=1}^K \hat{\rho}^{(k)}(\boldsymbol{x}, \mu) \\ \text{s.t.} & \mathcal{A}\boldsymbol{x} = \boldsymbol{b}, \\ & \boldsymbol{x} \in \mathcal{G}, \boldsymbol{x} \succ \boldsymbol{0}, \end{array} \quad \& \quad \begin{array}{ll} \min & \hat{\rho}^{(k)}(\boldsymbol{x}, \mu) := \boldsymbol{d}^{(k)} \bullet \boldsymbol{y}^{(k)} - \mu \ln\det \boldsymbol{y}^{(k)} \\ \text{s.t.} & \mathcal{W}^{(k)}\boldsymbol{y}^{(k)} = \boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x}, \\ & \boldsymbol{y}^{(k)} \in \hat{\mathcal{G}}, \boldsymbol{y}^{(k)} \succ \boldsymbol{0}. \end{array} \tag{6}$$

We define $\sum_{k=1}^{K} \hat{\rho}^{(k)}(\mu, \boldsymbol{x}) := \infty$ when the second-stage problem in (6) is infeasible for some k. The Lagrangian dual of the second-stage problem in (6) is the problem

$$\begin{array}{ll} \max & \left(\boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x}\right)^{\mathsf{T}} \boldsymbol{z}^{(k)} + \mu \ln\det \boldsymbol{s}^{(k)} \\ \text{s.t.} & \mathcal{W}^{(k)\dagger}\boldsymbol{z}^{(k)} + \boldsymbol{s}^{(k)} = \boldsymbol{d}^{(k)}, \\ & \boldsymbol{s}^{(k)} \in \hat{\mathcal{G}}^{\perp}, \boldsymbol{s}^{(k)} \succ \boldsymbol{0}, \end{array} \tag{7}$$

which is the logarithmic barrier problem associated with Problem (5). As the second-stage problem in (6) and Problem (7) are, respectively, concave and convex, $\boldsymbol{y}^{(k)}$ and $(\boldsymbol{z}^{(k)}, \boldsymbol{s}^{(k)})$ are optimal solutions to the second-stage problem in (6) and Problem (7), respectively, iff they satisfy the following optimality conditions:

$$\begin{array}{l} \boldsymbol{y}^{(k)} \circ \boldsymbol{s}^{(k)} = \mu\, \hat{\boldsymbol{e}}, \\ \mathcal{W}^{(k)}\boldsymbol{y}^{(k)} = \boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x},\ \boldsymbol{y}^{(k)} \succ \boldsymbol{0}, \\ \mathcal{W}^{(k)\dagger}\boldsymbol{z}^{(k)} + \boldsymbol{s}^{(k)} = \boldsymbol{d}^{(k)},\ \boldsymbol{s}^{(k)} \succ \boldsymbol{0}. \end{array} \tag{8}$$

Later, in this section, we need to compute the Fréchet derivatives of the recourse function. This in turn requires differentiating the system (8) and solving the resulting system. In fact, due to the fact that $\boldsymbol{y}^{(k)}$ and $\boldsymbol{s}^{(k)}$ may not operator commute (see Appendix A for definition), the resulting system may not have unique solution. We can overcome this difficulty by scaling the variables used in System (8) in a way that the scaled variables operator commute, or equivalently, the scaled elements are simultaneously decomposed (see Theorem 8). We use an effective way of scaling used by Schmieta and Alizadeh [20] for (finite-dimensional) symmetric cone programming.

Let $\boldsymbol{p} \in \hat{\mathcal{S}}$ be such that $\boldsymbol{p} \succ \boldsymbol{0}$. Note that $\mathcal{Q}_{\boldsymbol{p}}\mathcal{Q}_{\boldsymbol{p}^{-1}} = \mathcal{Q}_{\boldsymbol{p}}(\mathcal{Q}_{\boldsymbol{p}})^{-1} = \hat{\mathcal{I}}$, where $\mathcal{Q}_{\boldsymbol{p}}$ is the quadratic representation of $\boldsymbol{p}$ defined in Section 2. From now on, with respect to $\boldsymbol{p} \succ \boldsymbol{0}$ and for each $k = 1, 2, \ldots, K$, we define

$$\widetilde{\boldsymbol{y}}^{(k)} := \mathcal{Q}_{\boldsymbol{p}}\boldsymbol{y}^{(k)}, \widetilde{\boldsymbol{s}}^{(k)} := \mathcal{Q}_{\boldsymbol{p}^{-1}}\boldsymbol{s}^{(k)}, \widetilde{\boldsymbol{d}}^{(k)} := \mathcal{Q}_{\boldsymbol{p}^{-1}}\boldsymbol{d}^{(k)}, \text{ and } \widetilde{\boldsymbol{w}}_j^{(k)} := \boldsymbol{w}_j^{(k)}\mathcal{Q}_{\boldsymbol{p}^{-1}},$$

and define the map $\widetilde{\mathcal{W}}^{(k)} : \hat{\mathcal{S}} \longrightarrow \mathbb{R}^{m_2}$ by

$$\widetilde{\mathcal{W}}^{(k)}\widetilde{\boldsymbol{y}}^{(k)} := \begin{bmatrix} \widetilde{\boldsymbol{w}}_1^{(k)} \bullet \widetilde{\boldsymbol{y}}^{(k)} \\ \widetilde{\boldsymbol{w}}_2^{(k)} \bullet \widetilde{\boldsymbol{y}}^{(k)} \\ \vdots \\ \widetilde{\boldsymbol{w}}_{m_2}^{(k)} \bullet \widetilde{\boldsymbol{y}}^{(k)} \end{bmatrix}.$$

The following proposition is immediate.

Proposition 1 *$\left(\boldsymbol{y}^{(k)}, \boldsymbol{s}^{(k)}\right)$ satisfies the optimality conditions (8) iff $\left(\widetilde{\boldsymbol{y}}^{(k)}, \widetilde{\boldsymbol{s}}^{(k)}\right)$ satisfies the relaxed optimality conditions:*

$$\begin{array}{l} \widetilde{\boldsymbol{y}}^{(k)} \circ \widetilde{\boldsymbol{s}}^{(k)} = \mu\, \hat{\boldsymbol{e}}, \\ \widetilde{\mathcal{W}}^{(k)}\widetilde{\boldsymbol{y}}^{(k)} = \boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x},\ \widetilde{\boldsymbol{y}}^{(k)} \succ \boldsymbol{0}, \\ \widetilde{\mathcal{W}}^{(k)\dagger}\boldsymbol{z}^{(k)} + \widetilde{\boldsymbol{s}}^{(k)} = \widetilde{\boldsymbol{d}}^{(k)},\ \widetilde{\boldsymbol{s}}^{(k)} \succ \boldsymbol{0}. \end{array} \tag{9}$$

Proof The proof follows from Lemma 3, and the fact that $\mathcal{Q}_{\boldsymbol{p}}(\hat{\mathcal{S}}^{+}) = \hat{\mathcal{S}}^{+}$, and likewise, as an operator, $\mathcal{Q}_{\boldsymbol{p}}(\text{int}\,\hat{\mathcal{S}}^{+}) = \text{int}\,\hat{\mathcal{S}}^{+}$, because $\hat{\mathcal{S}}^{+}$ is a symmetric cone as shown in Corollary 1.

After this transformation, Problem (6) becomes

$$\begin{array}{llcll} \min & \eta(\mu, \boldsymbol{x}) := \boldsymbol{c} \bullet \boldsymbol{x} - \mu \ln\det \boldsymbol{x} + \sum_{k=1}^{K} \rho^{(k)}(\boldsymbol{x}, \mu) & \& & \min & \rho^{(k)}(\boldsymbol{x}, \mu) := \widetilde{\boldsymbol{d}}^{(k)} \bullet \widetilde{\boldsymbol{y}}^{(k)} - \mu \ln\det \widetilde{\boldsymbol{y}}^{(k)} \\ \text{s.t.} & \mathcal{A}\boldsymbol{x} = \boldsymbol{b}, & & \text{s.t.} & \widetilde{\mathcal{W}}^{(k)}\widetilde{\boldsymbol{y}}^{(k)} = \boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x}, \\ & \boldsymbol{x} \in \mathcal{G}, \boldsymbol{x} \succ \boldsymbol{0}, & & & \widetilde{\boldsymbol{y}}^{(k)} \in \hat{\mathcal{G}}, \widetilde{\boldsymbol{y}}^{(k)} \succ \boldsymbol{0}, \end{array} \tag{10}$$

and Problem (7) becomes

$$\begin{array}{ll} \max & \left(\boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x}\right)^{\mathsf{T}} \boldsymbol{z}^{(k)} + \mu \ln\det \widetilde{\boldsymbol{s}}^{(k)} \\ \text{s.t.} & \widetilde{\mathcal{W}}^{(k)\dagger}\boldsymbol{z}^{(k)} + \widetilde{\boldsymbol{s}}^{(k)} = \widetilde{\boldsymbol{d}}^{(k)}, \\ & \widetilde{\boldsymbol{s}}^{(k)} \in \hat{\mathcal{G}}^{\perp}, \widetilde{\boldsymbol{s}}^{(k)} \succ \boldsymbol{0}. \end{array} \tag{11}$$

The function $\rho^{(k)}(\mu, \boldsymbol{x})$ is called the *barrier recourse function* and the function $\eta(\mu, \boldsymbol{x})$ is called the *composite barrier function*. Note that the second-stage problems in (6) and (10) have the same minimizer but their optimal objective values are equal up to the constant $2\mu \ln\det \boldsymbol{p}$ (as $\det(\mathcal{Q}_{\boldsymbol{p}}\boldsymbol{q}) = \det^2(\boldsymbol{p})\det(\boldsymbol{q})$ for any two elements $\boldsymbol{p}, \boldsymbol{q} \in \hat{\mathcal{S}}$ [8, Proposition 16]). Similarly, Problems (7) and (11) have the same maximizer but their optimal objective values differ by the constant $2\mu \ln\det \boldsymbol{p}^{-1}$.

Throughout the rest of the paper, for a given $\mu > 0$, we denote the optimal solution of the first-stage problem (10) by $\boldsymbol{x}(\mu)$, or simply by $\boldsymbol{x}$, and the solutions of the optimality conditions (9) by $(\widetilde{\boldsymbol{y}}^{(k)}(\mu, \boldsymbol{x}), \boldsymbol{z}^{(k)}(\mu, \boldsymbol{x}), \widetilde{\boldsymbol{s}}^{(k)}(\mu, \boldsymbol{x}))$, or simply by $(\widetilde{\boldsymbol{y}}^{(k)}, \boldsymbol{z}^{(k)}, \widetilde{\boldsymbol{s}}^{(k)})$.

Problem (10) can be written equivalently as the large-scale deterministic program:

$$\begin{array}{ll} \min & \eta(\mu, \boldsymbol{x}) := \boldsymbol{c} \bullet \boldsymbol{x} - \mu \ln\det \boldsymbol{x} + \displaystyle\sum_{k=1}^{K} \left(\widetilde{\boldsymbol{d}}^{(k)} \bullet \widetilde{\boldsymbol{y}}^{(k)} - \mu \ln\det \widetilde{\boldsymbol{y}}^{(k)} \right) \\ \text{s.t.} & \mathcal{A}\boldsymbol{x} = \boldsymbol{b}, \\ & \widetilde{\mathcal{W}}^{(k)} \widetilde{\boldsymbol{y}}^{(k)} = \boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x},\ k = 1, 2, \ldots, K, \\ & \boldsymbol{x} \in \mathcal{G},\ \widetilde{\boldsymbol{y}}^{(k)} \in \hat{\mathcal{G}},\ k = 1, 2, \ldots, K, \\ & \boldsymbol{x} \succ \boldsymbol{0},\ \widetilde{\boldsymbol{y}}^{(k)} \succ \boldsymbol{0},\ k = 1, 2, \ldots, K. \end{array} \tag{12}$$

Each choice of $\boldsymbol{p}$ leads us to a different search direction. As we indicated earlier, our focus is on the following set of scalings $C(\boldsymbol{y}^{(k)}, \boldsymbol{s}^{(k)}) := \{\boldsymbol{p} \succ \boldsymbol{0} \ :\ \widetilde{\boldsymbol{y}}^{(k)} \text{ and } \widetilde{\boldsymbol{s}}^{(k)} \text{ operator commute}\}$. The set of directions $\Delta\boldsymbol{x}$ arising from those $\boldsymbol{p}$ in $C(\boldsymbol{y}^{(k)}, \boldsymbol{s}^{(k)})$ is called the *commutative class of directions*, and a direction in this class is called a *commutative direction*.

Clearly $\boldsymbol{p} = \hat{\boldsymbol{e}}$ may not be in $C(\boldsymbol{y}^{(k)}, \boldsymbol{s}^{(k)})$. Analogue to the discussion in [20, Section 3], we focus on three choices of $\boldsymbol{p}$ that guarantee that $C(\boldsymbol{y}^{(k)}, \boldsymbol{s}^{(k)}) \neq \emptyset$: Firstly, we choose $\boldsymbol{p} = {\boldsymbol{y}^{(k)}}^{1/2}$ to get $\widetilde{\boldsymbol{y}}^{(k)} = \hat{\boldsymbol{e}}$. Secondly, we choose $\boldsymbol{p} = {\boldsymbol{s}^{(k)}}^{-1/2}$ to get $\widetilde{\boldsymbol{s}}^{(k)} = \hat{\boldsymbol{e}}$. These two choices of directions are well-known in commutative classes and form a class of Newton directions derived by Helmberg et. al [22], Monteiro [23], and Kojima et. al [24], and referred to as the HRVW/KSH/M directions.

Thirdly, we choose

$$\boldsymbol{p} = \left(\mathcal{Q}_{\boldsymbol{s}^{(k)1/2}} \left(\mathcal{Q}_{\boldsymbol{s}^{(k)1/2}} \boldsymbol{y}^{(k)} \right)^{-1/2} \right)^{-1/2} = \left(\mathcal{Q}_{\boldsymbol{y}^{(k)-1/2}} \left(\mathcal{Q}_{\boldsymbol{y}^{(k)1/2}} \boldsymbol{s}^{(k)} \right)^{1/2} \right)^{-1/2}$$

to get $\widetilde{\boldsymbol{s}}^{(k)} = \mathcal{Q}_{\boldsymbol{p}} \boldsymbol{s}^{(k)} = \mathcal{Q}_{\boldsymbol{p}^{-1}} \boldsymbol{y}^{(k)} = \widetilde{\boldsymbol{y}}^{(k)}$ (as $Q_p^2 \boldsymbol{s}^{(k)} = \boldsymbol{y}^{(k)}$). This choice leads to the NT direction (due to Nesterov and Todd).

Now, we define the following feasibility sets:

$$\begin{array}{lcl} \mathcal{F}_1 & := & \left\{ \boldsymbol{x} \in \mathcal{G} : \mathcal{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \succ \boldsymbol{0} \right\}; \\ \mathcal{F}^{(k)}(\boldsymbol{x}) & := & \left\{ \widetilde{\boldsymbol{y}}^{(k)} \in \hat{\mathcal{G}} : \widetilde{\mathcal{W}}^{(k)} \widetilde{\boldsymbol{y}}^{(k)} = \boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x}, \widetilde{\boldsymbol{y}}^{(k)} \succ \boldsymbol{0} \right\} \text{ for } k = 1, 2, \ldots, K; \\ \mathcal{F}_2^{(k)} & := & \left\{ \boldsymbol{x} \in \mathcal{G} : \mathcal{F}^{(k)}(\boldsymbol{x}) \neq \emptyset \right\} \text{ for } k = 1, 2, \ldots, K; \\ \mathcal{F}_2 & := & \bigcap_{k=1}^{K} \mathcal{F}_2^{(k)}; \\ \mathcal{F}_0 & := & \mathcal{F}_1 \bigcap \mathcal{F}_2; \\ \mathcal{F} & := & \Big\{ (\boldsymbol{x}, \boldsymbol{\lambda}), \left(\widetilde{\boldsymbol{y}}^{(1)}, \boldsymbol{z}^{(1)}, \ldots, \widetilde{\boldsymbol{y}}^{(K)}, \boldsymbol{z}^{(K)} \right) : \mathcal{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \succ \boldsymbol{0}, \sum_{k=1}^{K} \mathcal{T}^{(k)\dagger} \boldsymbol{z}^{(k)} \prec \boldsymbol{c}, \boldsymbol{x} \bullet (\boldsymbol{c} - \mathcal{A}^{\dagger}\boldsymbol{\lambda}) = 0, \\ & & \widetilde{\mathcal{W}}^{(k)} \widetilde{\boldsymbol{y}}^{(k)} = \boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x}, \widetilde{\boldsymbol{y}}^{(k)} \succ \boldsymbol{0}, \widetilde{\mathcal{W}}^{(k)\dagger} \boldsymbol{z}^{(k)} \prec \widetilde{\boldsymbol{d}}^{(k)}, \widetilde{\boldsymbol{y}}^{(k)} \bullet (\widetilde{\boldsymbol{d}}^{(k)} - \widetilde{\mathcal{W}}^{(k)\dagger} \boldsymbol{z}^{(k)}) = 0, k = 1, 2, \ldots, K \Big\}. \end{array}$$

Here $\boldsymbol{\lambda} \in \mathbb{R}^{m_1}$ is the first-stage dual multiplier. The, we make the following two assumptions:

Assumption 1 *The elements $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{m_1}$ are linearly independent in $\mathcal{S}$, and for $k = 1, 2, \ldots, K$, the elements $\boldsymbol{t}_1^{(k)}, \boldsymbol{t}_2^{(k)}, \ldots, \boldsymbol{t}_{m_2}^{(k)}$ are linearly independent in $\mathcal{S}$, and the elements $\boldsymbol{w}_1^{(k)}, \boldsymbol{w}_2^{(k)}, \ldots, \boldsymbol{w}_{m_2}^{(k)}$ are linearly independent in $\hat{\mathcal{S}}$.*

Assumption 2 *The feasibility set $\mathcal{F}$ is nonempty.*

Assumption 1 is necessary to ensure invertibility of operators in $\mathcal{S}$ and $\hat{\mathcal{S}}$. Assumption 2 is necessary to guarantee strong duality for the first- and second-stage problems. Therefore, it implies that Problems

(10-12) have a unique solutions. Note that for a given $\mu > 0$, $\sum_{k=1}^{K} \rho^{(k)}(\mu, \boldsymbol{x}) < \infty$ iff $\boldsymbol{x} \in \mathcal{F}_2$. So, the feasible region for (10) is described implicitly by $\mathcal{F}_0$.

The optimal solutions of Problem (10) and those of Problem (12) have the following relationship. The point $(\boldsymbol{x}(\mu); \widetilde{\boldsymbol{y}}^{(1)}(\mu); \cdots; \widetilde{\boldsymbol{y}}^{(K)}(\mu))$ is the optimal solution of (12) iff $\boldsymbol{x}(\mu)$, $(\widetilde{\boldsymbol{y}}^{(1)}(\mu); \cdots; \widetilde{\boldsymbol{y}}^{(K)}(\mu))$ are the optimal solutions for (10) for given μ and $\boldsymbol{x}$.

4 Fréchet derivatives

In this section, we compete $\mathcal{D}_{x}\eta(\mu, \boldsymbol{x})$ and $\mathcal{D}^2_{xx}\eta(\mu, \boldsymbol{x})$ needed for determining the Newton direction $\Delta \boldsymbol{x}$. First, we compute the first and second Fréchet derivatives of the function $\ln\det(\boldsymbol{x})$. We have

$$
\begin{array}{rcl}
\mathcal{D}_{x} (\ln\det \boldsymbol{x}) (\boldsymbol{\xi}) & = & \{\ln\det (\boldsymbol{x} + t\boldsymbol{\xi})\}'|_{t=0} \\
& = & \{\ln\det (\mathcal{Q}_{\boldsymbol{x}^{1/2}} (\boldsymbol{e} + t\mathcal{Q}_{\boldsymbol{x}^{-1/2}}\boldsymbol{\xi}))\}'|_{t=0} \\
& = & \{\ln\det (\boldsymbol{x}) + \ln\det (\boldsymbol{e} + t\mathcal{Q}_{\boldsymbol{x}^{-1/2}}\boldsymbol{\xi}))\}'|_{t=0} \\
& = & \operatorname{trace} (\mathcal{Q}_{\boldsymbol{x}^{-1/2}}\boldsymbol{\xi}) \\
& = & \operatorname{trace}(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}) = \boldsymbol{x}^{-1} \bullet \boldsymbol{\xi},
\end{array}
$$

$$
\begin{array}{rcl}
\mathcal{D}^2_{xx} (\ln\det \boldsymbol{x}) (\boldsymbol{\xi}, \boldsymbol{\xi}) & = & \mathcal{D}_{x}(\mathcal{D}_{x} (\ln\det \boldsymbol{x}) (\boldsymbol{\xi}))(\boldsymbol{\xi}) \\
& = & \mathcal{D}_{x}(\boldsymbol{x}^{-1} \bullet \boldsymbol{\xi})(\boldsymbol{\xi}) \\
& = & \left\{\left((\boldsymbol{x} + t\boldsymbol{\xi})^{-1} \bullet \boldsymbol{\xi}\right)\right\}'\Big|_{t=0} \\
& = & -\left(\boldsymbol{x}^{-2} \circ \boldsymbol{\xi}\right) \bullet \boldsymbol{\xi} = -\left(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}\right) \bullet \left(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}\right),
\end{array} \tag{13}
$$

$$
\begin{array}{rcl}
\mathcal{D}^3_{xxx} (\ln\det \boldsymbol{x}) (\boldsymbol{\xi}, \boldsymbol{\xi}, \boldsymbol{\xi}) & = & \mathcal{D}_{x} \left(\mathcal{D}^2_{xx} (\ln\det \boldsymbol{x}) (\boldsymbol{\xi}, \boldsymbol{\xi})\right) (\boldsymbol{\xi}) \\
& = & -\mathcal{D}_{x} \left(\left(\boldsymbol{x}^{-2} \circ \boldsymbol{\xi}\right) \bullet \boldsymbol{\xi}\right) (\boldsymbol{\xi}) \\
& = & -\left\{\left((\boldsymbol{x} + t\boldsymbol{\xi})^{-2} \circ \boldsymbol{\xi}\right) \bullet \boldsymbol{\xi}\right\}'\Big|_{t=0} \\
& = & 2\left(\left(\boldsymbol{x}^{-3} \circ \boldsymbol{\xi}\right) \circ \boldsymbol{\xi}\right) \bullet \boldsymbol{\xi} = 2\left(\left(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}\right) \circ \left(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}\right)\right) \bullet \left(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}\right).
\end{array} \tag{14}
$$

In summary, the first and second Fréchet derivatives of the function $\ln\det(\boldsymbol{x})$ read

$$
\mathcal{D}_{x} \ln\det(\boldsymbol{x}) = 2\boldsymbol{x}^{-1}, \quad \text{and} \quad \mathcal{D}^2_{xx} \ln\det(\boldsymbol{x}) = -2\mathcal{Q}_{\boldsymbol{x}^{-1}}.
$$

Second, we compute the first and second Fréchet derivatives of the function $\rho^{(k)}(\mu, \boldsymbol{x})$. Note that

$$
\rho^{(k)}(\mu, \boldsymbol{x}) = \widetilde{\boldsymbol{d}}^{(k)} \bullet \widetilde{\boldsymbol{y}}^{(k)} - \mu \ln\det \widetilde{\boldsymbol{y}}^{(k)} = \widetilde{\boldsymbol{y}}^{(k)} \bullet \left(\widetilde{\boldsymbol{d}}^{(k)} - \widetilde{\boldsymbol{s}}^{(k)}\right) + \widetilde{\boldsymbol{y}}^{(k)} \bullet \widetilde{\boldsymbol{s}}^{(k)} - \mu \ln\det \widetilde{\boldsymbol{y}}^{(k)} = \widetilde{\boldsymbol{y}}^{(k)} \bullet \left(\widetilde{\boldsymbol{d}}^{(k)} - \widetilde{\boldsymbol{s}}^{(k)}\right) - \mu \ln\det \widetilde{\boldsymbol{y}}^{(k)},
$$

where the last equality follows from the fact that $\widetilde{\boldsymbol{y}}^{(k)} \in \hat{\mathcal{G}}$ and $\widetilde{\boldsymbol{s}}^{(k)} \in \hat{\mathcal{G}}^{\perp}$.

From (9), we have

$$
\widetilde{\boldsymbol{y}}^{(k)} \bullet \left(\widetilde{\boldsymbol{d}}^{(k)} - \boldsymbol{s}^{(k)}\right) = \widetilde{\boldsymbol{y}}^{(k)} \bullet \widetilde{\mathcal{W}}^{(k)\dagger} \boldsymbol{z}^{(k)} = \sum_{i=1}^{m_2} z_i^{(k)} \left(\boldsymbol{w}_j^{(k)} \bullet \widetilde{\boldsymbol{y}}^{(k)}\right) = \left(\widetilde{\mathcal{W}}^{(k)} \widetilde{\boldsymbol{y}}^{(k)}\right)^{\mathsf{T}} \boldsymbol{z}^{(k)} = \left(\boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x}\right)^{\mathsf{T}} \boldsymbol{z}^{(k)}.
$$

We also have

$$
\ln\det \widetilde{\boldsymbol{y}}^{(k)} + \ln\det \widetilde{\boldsymbol{s}}^{(k)} = \ln\left(\det \widetilde{\boldsymbol{y}}^{(k)} \det \widetilde{\boldsymbol{s}}^{(k)}\right) = \ln\det\left(\widetilde{\boldsymbol{y}}^{(k)} \circ \widetilde{\boldsymbol{s}}^{(k)}\right) = \ln\det(\mu\hat{\boldsymbol{e}}) = \ln \mu^{2r_2} = 2r_2 \ln \mu.
$$

Thus

$$
\rho^{(k)}(\mu, \boldsymbol{x}) = \left(\boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x}\right)^{\mathsf{T}} \boldsymbol{z}^{(k)} + \mu \ln\det \widetilde{\boldsymbol{s}}^{(k)} - 2r_2\mu \ln \mu, \tag{15}
$$

which is the optimal objective of Problem (11).

To compute $\mathcal{D}_{x}\eta(\mu, \boldsymbol{x})$ and $\mathcal{D}^2_{xx}\eta(\mu, \boldsymbol{x})$, we need to compute $\mathcal{D}_{x}\boldsymbol{z}^{(k)}$; the first Fréchet derivative of $\boldsymbol{z}^{(k)}$ with respect to $\boldsymbol{x}$. Differentiating (9) with respect to $\boldsymbol{x}$, we get the system

$$
\begin{array}{l}
\mathcal{D}_{x}\widetilde{\boldsymbol{y}}^{(k)} = -\mu \, \mathcal{Q}_{\widetilde{\boldsymbol{s}}^{(k)-1}} \mathcal{D}_{x}\widetilde{\boldsymbol{s}}^{(k)}, \\
\widetilde{\mathcal{W}}^{(k)} \mathcal{D}_{x}\widetilde{\boldsymbol{y}}^{(k)} = -\mathcal{T}^{(k)}, \\
\widetilde{\mathcal{W}}^{(k)\dagger} \mathcal{D}_{x}\boldsymbol{z}^{(k)} + \mathcal{D}_{x}\widetilde{\boldsymbol{s}}^{(k)} = \boldsymbol{0}.
\end{array} \tag{16}
$$

Solving the system (16), we obtain

$$\begin{array}{rl} \mathcal{D}_x \boldsymbol{z}^{(k)} &= -R^{(k)^{-1}}\mathcal{T}^{(k)}, \\ \mathcal{D}_x \widetilde{\boldsymbol{y}}^{(k)} &= -\mathcal{P}^{(k)2}\widetilde{\mathcal{W}}^{(k)\dagger}R^{(k)^{-1}}\mathcal{T}^{(k)}, \\ \mathcal{D}_x \widetilde{\boldsymbol{s}}^{(k)} &= \widetilde{\mathcal{W}}^{(k)\dagger}R^{(k)^{-1}}\mathcal{T}^{(k)}, \end{array} \tag{17}$$

where, for $k = 1, 2, \ldots, K$, the maps $\mathcal{P}^{(k)} : \hat{\mathcal{S}} \longrightarrow \hat{\mathcal{S}}$ and $R^{(k)} : \mathbb{R}^{m_2} \longrightarrow \mathbb{R}^{m_2}$ are respectively defined by

$$\mathcal{P}^{(k)} := \mathcal{P}^{(k)}(\mu, \boldsymbol{x}) := \sqrt{\mu}\ \mathcal{Q}^{-1/2}_{\widetilde{\boldsymbol{s}}^{(k)}}, \quad \text{and} \quad R^{(k)} := R^{(k)}(\mu, \boldsymbol{x}) := \widetilde{\mathcal{W}}^{(k)}\mathcal{P}^{(k)^2}\widetilde{\mathcal{W}}^{(k)\dagger}. \tag{18}$$

Now, by differentiating (15) and using the optimality conditions (9) and (17), we can conclude that the first and second Fréchet derivatives of $\rho^{(k)}(\mu, \boldsymbol{x})$ read

$$\mathcal{D}_x \rho^{(k)}(\mu, \boldsymbol{x}) = -\mathcal{T}^{(k)\dagger}\boldsymbol{z}^{(k)}(\mu, \boldsymbol{x}), \quad \text{and} \quad \mathcal{D}^2_{xx}\rho^{(k)}(\mu, \boldsymbol{x}) = -\mathcal{T}^{(k)\dagger}\mathcal{D}_x \boldsymbol{z}^{(k)}(\mu, \boldsymbol{x}) = -\mathcal{T}^{(k)\dagger}R^{(k)^{-1}}\mathcal{T}^{(k)}.$$

Finally, using (17), the first and second Fréchet derivatives of the composite barrier function $\eta(\mu, \boldsymbol{x})$ read

$$\begin{array}{l} \mathcal{D}_x \eta(\mu, \boldsymbol{x}) = \boldsymbol{c} - \mu\, \boldsymbol{x}^{-1} + \displaystyle\sum_{k=1}^{K} \mathcal{D}_x \rho^{(k)}(\mu, \boldsymbol{x}) = \boldsymbol{c} - \mu\, \boldsymbol{x}^{-1} - \sum_{k=1}^{K} \mathcal{T}^{(k)\dagger}\boldsymbol{z}^{(k)}(\mu, \boldsymbol{x}), \\ \mathcal{D}^2_{xx}\eta(\mu, \boldsymbol{x}) = \mu\, \mathcal{Q}_{\boldsymbol{x}^{-1}} - \displaystyle\sum_{k=1}^{K} \mathcal{T}^{(k)\dagger}\mathcal{D}_x \boldsymbol{z}^{(k)}(\mu, \boldsymbol{x}) = \mu\, \mathcal{Q}_{\boldsymbol{x}^{-1}} - \sum_{k=1}^{K} \mathcal{T}^{(k)\dagger}R^{(k)^{-1}}\mathcal{T}^{(k)}. \end{array} \tag{19}$$

For our subsequent development, we also need to compute the partial derivatives of $\mathcal{D}_x\eta(\mu, \boldsymbol{x})$ and $\mathcal{D}^2_{xx}\eta(\mu, \boldsymbol{x})$ with respect to μ. Let $(\widetilde{\boldsymbol{y}}^{(k)'}, \boldsymbol{z}^{(k)'}, \widetilde{\boldsymbol{s}}^{(k)'})$ denote the partial derivatives of $(\widetilde{\boldsymbol{y}}^{(k)}(\mu, \boldsymbol{x}), \boldsymbol{z}^{(k)}(\mu, \boldsymbol{x}), \widetilde{\boldsymbol{s}}^{(k)}(\mu, \boldsymbol{x}))$ with respect to μ. By differentiating (9) with respect to μ, we get the system

$$\begin{array}{rl} \widetilde{\boldsymbol{y}}^{(k)'} &= -\mu\, \mathcal{Q}_{\widetilde{\boldsymbol{s}}^{(k)^{-1}}}\widetilde{\boldsymbol{s}}^{(k)'} + \widetilde{\boldsymbol{s}}^{(k)^{-1}}, \\ \widetilde{\mathcal{W}}^{(k)}\boldsymbol{y}^{(k)'} &= \boldsymbol{0}, \\ \widetilde{\mathcal{W}}^{(k)\dagger}\boldsymbol{z}^{(k)'} + \widetilde{\boldsymbol{s}}^{(k)'} &= \boldsymbol{0}. \end{array} \tag{20}$$

Solving the system (20), we obtain

$$\begin{array}{rl} \boldsymbol{z}^{(k)'} &= -R^{(k)^{-1}}\widetilde{\mathcal{W}}^{(k)}\mathcal{P}^{(k)^2}\widetilde{\boldsymbol{y}}^{(k)^{-1}}, \\ \widetilde{\boldsymbol{y}}^{(k)'} &= \mathcal{P}^{(k)^2}\left(\hat{\mathcal{I}} - \widetilde{\mathcal{W}}^{(k)\dagger}R^{(k)^{-1}}\widetilde{\mathcal{W}}^{(k)}\mathcal{P}^{(k)^2}\right)\widetilde{\boldsymbol{y}}^{(k)^{-1}}, \\ \widetilde{\boldsymbol{s}}^{(k)'} &= \widetilde{\mathcal{W}}^{(k)\dagger}R^{(k)^{-1}}\widetilde{\mathcal{W}}^{(k)}\mathcal{P}^{(k)^2}\widetilde{\boldsymbol{y}}^{(k)^{-1}}. \end{array} \tag{21}$$

By differentiating (19) with respect to μ and using (21), we get

$$\begin{array}{l} \{\mathcal{D}_x\eta(\mu, \boldsymbol{x})\}' = -\boldsymbol{x}^{-1} - \displaystyle\sum_{k=1}^{K} \mathcal{T}^{(k)\dagger}\, \widetilde{\boldsymbol{z}}^{(k)'} = -\boldsymbol{x}^{-1} + \sum_{k=1}^{K} \mathcal{T}^{(k)\dagger}R^{(k)^{-1}}\widetilde{\mathcal{W}}^{(k)}\mathcal{P}^{(k)^2}\widetilde{\boldsymbol{y}}^{(k)^{-1}}, \\ \left\{\mathcal{D}^2_{xx}\eta(\mu, \boldsymbol{x})\right\}' = \mathcal{Q}_{\boldsymbol{x}^{-1}} - \displaystyle\sum_{k=1}^{K} \mathcal{T}^{(k)\dagger}R^{(k)^{-1}}R^{(k)'}R^{(k)^{-1}}\mathcal{T}^{(k)} = \mathcal{Q}_{\boldsymbol{x}^{-1}} + 2\sum_{k=1}^{K} \mathcal{T}^{(k)\dagger}R^{(k)^{-1}}\widetilde{\mathcal{W}}^{(k)}\mathcal{P}^{(k)}\left\{\mathcal{P}^{(k)}\right\}'\widetilde{\mathcal{W}}^{(k)\dagger}R^{(k)^{-1}}\mathcal{T}^{(k)}. \end{array} \tag{22}$$

5 Self-concordance

In this section, we present two fundamental properties of the recourse function $\eta(\mu, \cdot)$. In particular, we show that the barrier function is a strongly self-concordant barrier and that the set of barrier functions for positive values of barrier parameters comprises a self-concordant family. These functions are at the heart of the decomposition interior-point methods that we develop. The results obtained in this section are of fundamental importance because they lead us to prove the complexity results of the proposed algorithms.

To show that $\eta(\mu,\cdot)$ is a μ strongly self-concordant barrier on $\mathcal{F}_0$, we first introduce the definition of a self-concordant function. Nesterov and Nemirovskii introduced this definition for finite-dimensional optimization (see Definition 2.1.1 in [25]). The definition is introduced below for optimization in a Hilbert space (see also [9]). Throughout the rest of the paper, we use $\mathbb{R}_{++}$ to denote the set of all positive real numbers.

Definition 1 *Let C be an open nonempty convex subset of a Hilbert space $\mathcal{H}$, and let f be thrice Fréchet differentiable, convex mapping from C to $\mathbb{R}$. Then f is called α-*self-concordant *on C with the parameter $\alpha > 0$ if for every $\boldsymbol{x} \in C$ and $\boldsymbol{\xi} \in \mathcal{H}$, the following inequality holds*

$$\left|\mathcal{D}^3_{xxx} f(x)(\xi,\xi,\xi)\right| \quad \leq \quad 2\,\alpha^{-1/2}\left(\mathcal{D}^2_{xx} f(x)(\xi,\xi)\right)^{3/2}. \tag{23}$$

An α-self-concordant function f on C is called strongly α-self-concordant *if f tends to infinity for any sequence approaching a boundary point of C. The parameter α is called the* complexity value *of the self-concordant function f.*

The proof of self-concordance of the recourse function depends on the following intermediate lemma.

Lemma 1 Let $\mu > 0$ and $\boldsymbol{x} \in \mathcal{F}_2^{(k)}$, and define the univariate function $\Xi^{(k)}(t) := \mathcal{D}^2_{xx}\rho^{(k)}(\mu, \boldsymbol{x}+t\boldsymbol{\xi})(\boldsymbol{\xi},\boldsymbol{\xi})$ for $\boldsymbol{\xi} \in \mathcal{S}$. Then

$$|\Xi^{(k)\prime}(0)| \leq 2\mu^{-\frac{1}{2}}|\Xi^{(k)}(0)|^{\frac{3}{2}}.$$

Proof Let $\left(\widetilde{\boldsymbol{y}}^{(k)}(t), z^{(k)}(t), \widetilde{\boldsymbol{s}}^{(k)}(t)\right) := \left(\widetilde{\boldsymbol{y}}^{(k)}(\mu, \boldsymbol{x}+t\boldsymbol{\xi}), z^{(k)}(\mu, \boldsymbol{x}+t\boldsymbol{\xi}), \widetilde{\boldsymbol{s}}^{(k)}(\mu, \boldsymbol{x}+t\boldsymbol{\xi})\right)$. Let also

$$\mathcal{P}^{(k)}(t) := \mathcal{P}^{(k)}(\mu, \boldsymbol{x}+t\boldsymbol{\xi}), \text{ and } R^{(k)}(t) := R^{(k)}(\mu, \boldsymbol{x}+t\boldsymbol{\xi}).$$

Using the notations introduced in Section 3, we have $\left(\widetilde{\boldsymbol{y}}^{(k)}, z^{(k)}, \widetilde{\boldsymbol{s}}^{(k)}\right) = \left(\widetilde{\boldsymbol{y}}^{(k)}(0), z^{(k)}(0), \widetilde{\boldsymbol{s}}^{(k)}(0)\right)$, $\mathcal{P}^{(k)} = \mathcal{P}^{(k)}(0)$, and $R^{(k)} = R^{(k)}(0)$. Let $\boldsymbol{\zeta} \in \mathcal{S}$. From (19), we have

$$\Xi^{(k)}(t) = \boldsymbol{\zeta} \bullet \mathcal{T}^{(k)\dagger} R^{(k)^{-1}}(t)\, \mathcal{T}^{(k)}\boldsymbol{\zeta}.$$

Note that $\{R^{(k)^{-1}}(t)\}' = -R^{(k)^{-1}}(t)\, R^{(k)\prime}(t)\, R^{(k)^{-1}}(t)$. Differentiating $\Xi^{(k)}(t)$ with respect to t, we get

$$|\Xi^{(k)\prime}(t)| \quad = \quad \left|\boldsymbol{\zeta} \bullet \mathcal{T}^{(k)\dagger} R^{(k)^{-1}}(t)\, \widetilde{\mathcal{W}}^{(k)} \mathcal{P}^{(k)^2}(t)\, \{\mathcal{P}^{(k)^{-2}}(t)\}' \mathcal{P}^{(k)^2}(t)\, \widetilde{\mathcal{W}}^{(k)\dagger} R^{(k)^{-1}}(t)\, \mathcal{T}^{(k)}\boldsymbol{\zeta}\right|. \tag{24}$$

We now intend to bound the term in the right-hand side of (24).

Note that $\mathcal{P}^{(k)}(t)\; \mathcal{Q}_{\widetilde{\boldsymbol{s}}^{(k)}(t)} \mathcal{P}^{(k)}(t) = \mu\hat{\mathcal{I}}$. Then, for any $\boldsymbol{\xi} \in \hat{\mathcal{S}}$, we have

$$\left|\boldsymbol{\xi} \bullet \mathcal{P}^{(k)}(t)\, \{\mathcal{P}^{(k)^{-2}}(t)\}' \mathcal{P}^{(k)}(t)\, \boldsymbol{\xi}\right| \leq \left\|\mathcal{Q}_{\widetilde{\boldsymbol{y}}^{(k)}(t)}\{\mathcal{P}^{(k)^{-2}}(t)\}'\right\|_2 \boldsymbol{\xi} \bullet \mathcal{P}^{(k)}(t)\, \mathcal{Q}_{\widetilde{\boldsymbol{y}}^{(k)^{-1}}(t)} \mathcal{P}^{(k)}(t)\, \boldsymbol{\xi} = \frac{1}{\mu}\left\|\mathcal{Q}_{\widetilde{\boldsymbol{y}}^{(k)}(t)}\{\mathcal{P}^{(k)^{-2}}(t)\}'\hat{\boldsymbol{e}}\right\|_2 \boldsymbol{\xi} \bullet \boldsymbol{\xi}. \tag{25}$$

Now,

$$\mathcal{Q}_{\widetilde{\boldsymbol{y}}^{(k)}(t)}\, \{\mathcal{P}^{(k)^{-2}}(t)\}'\hat{\boldsymbol{e}} = \frac{1}{\mu}\mathcal{Q}_{\widetilde{\boldsymbol{y}}^{(k)}(t)}\, \left(\mathcal{Q}_{\widetilde{\boldsymbol{s}}^{(k)}(t)+\widetilde{\boldsymbol{s}}^{(k)\prime}(t)} - \mathcal{Q}_{\widetilde{\boldsymbol{s}}^{(k)}(t)} - \mathcal{Q}_{\widetilde{\boldsymbol{s}}^{(k)\prime}(t)}\right)\hat{\boldsymbol{e}} = \widetilde{\boldsymbol{y}}^{(k)}(t) \circ \widetilde{\boldsymbol{s}}^{(k)\prime}(t) - \widetilde{\boldsymbol{s}}^{(k)}(t) \circ \widetilde{\boldsymbol{y}}^{(k)\prime}(t). \tag{26}$$

Using (17), we can write

$$\begin{aligned} \widetilde{\boldsymbol{y}}^{(k)\prime}(t) &= \mathcal{D}_x \widetilde{\boldsymbol{y}}^{(k)}(\mu, \boldsymbol{x}+t\boldsymbol{\zeta})(\boldsymbol{\zeta}) = -\mathcal{P}^{(k)^2}(t)\, \widetilde{\mathcal{W}}^{(k)\dagger} R^{(k)^{-1}}(t)\, \mathcal{T}^{(k)}\boldsymbol{\zeta}, \\ \widetilde{\boldsymbol{s}}^{(k)\prime}(t) &= \mathcal{D}_x \widetilde{\boldsymbol{s}}^{(k)}(\mu, \boldsymbol{x}+t\boldsymbol{\zeta})(\boldsymbol{\zeta}) = \widetilde{\mathcal{W}}^{(k)\dagger} R^{(k)^{-1}}(t)\, \mathcal{T}^{(k)}\boldsymbol{\zeta}. \end{aligned} \tag{27}$$

Using (26), (27), and the fact that $\mathcal{P}^{(k)}(t)\; \mathcal{Q}_{\widetilde{\boldsymbol{s}}^{(k)}(t)} \mathcal{P}^{(k)}(t) = \mu\hat{\mathcal{I}}$, we get

$$\left\|\mathcal{Q}_{\widetilde{\boldsymbol{y}}^{(k)}(t)}\{\mathcal{P}^{(k)^{-2}}(t)\}'\hat{\boldsymbol{e}}\right\|_2 \leq 2\left\|\mathcal{Q}_{\widetilde{\boldsymbol{s}}^{(k)^{1/2}}(t)}\; \mathcal{P}^{(k)^2}(t)\, \widetilde{\mathcal{W}}^{(k)\dagger} R^{(k)^{-1}}(t)\, \mathcal{T}^{(k)}\boldsymbol{\zeta}\right\|_F = 2^{3/2}\sqrt{\mu}\left(\boldsymbol{\zeta} \bullet \mathcal{T}^{(k)\dagger} R^{(k)^{-1}}(t) \mathcal{T}^{(k)}\boldsymbol{\zeta}\right)^{1/2}.$$

It then follows from the definition of $\Xi^{(k)}(t)$ that

$$\left\|\mathcal{Q}_{\widetilde{\boldsymbol{y}}^{(k)}(t)}\ \{\mathcal{P}^{(k)^{-2}}(t)\}'\right\|_2 \leq 2^{3/2}\ \sqrt{\mu}\ \Xi^{(k)^{1/2}}(t). \tag{28}$$

Combining (25) and (28), we conclude that for any $\boldsymbol{\xi} \in \mathcal{S}$, we have

$$|\boldsymbol{\xi} \bullet \mathcal{P}^{(k)}(t)\ \{\mathcal{P}^{(k)^{-2}}(t)\}'\mathcal{P}^{(k)}(t)\ \boldsymbol{\xi}| \leq \tfrac{2^{3/2}}{\sqrt{\mu}}\ \Xi^{(k)^{1/2}}(t)\ \boldsymbol{\xi} \bullet \boldsymbol{\xi}. \tag{29}$$

We set $t = 0$, take $\boldsymbol{\xi}^{(k)} := 2^{1/4}\ \mathcal{P}^{(k)}(0)\widetilde{\mathcal{W}}^{(k)\dagger}R^{(k)^{-1}}(0)\ \mathcal{T}^{(k)}\boldsymbol{\zeta}$ and use (24) and (29) to get finally

$$|\Xi^{(k)'}(0)| = \frac{1}{\sqrt{2}}\left|\boldsymbol{\xi}^{(k)} \bullet \mathcal{P}^{(k)}(t)\ \{\mathcal{P}^{(k)^{-2}}(t)\}'\mathcal{P}^{(k)}(t)\ \boldsymbol{\xi}^{(k)}\right| \leq \frac{2}{\sqrt{\mu}}\ \Xi^{(k)^{1/2}}(0)\ \boldsymbol{\xi}^{(k)} \bullet \boldsymbol{\xi}^{(k)} = \frac{2}{\sqrt{\mu}}\ \Xi^{(k)^{3/2}}(0).$$

The proof is complete.

We can now state and prove the following important theorem.

Theorem 1 *For any fixed $\mu > 0, \eta(\mu, \boldsymbol{x})$ is a μ strongly self-concordant barrier on $\mathcal{F}_0$.*

Proof It is defined that

$$\eta(\mu, \boldsymbol{x}) := \boldsymbol{c} \bullet \boldsymbol{x} - \mu \ln \det \boldsymbol{x} + \sum_{k=1}^{K} \rho^{(k)}(\boldsymbol{x}, \mu). \tag{30}$$

Let $\mu > 0$ be fixed and $\{\boldsymbol{x}_i\}_{i=1}^{\infty}$ be any sequence in $\mathcal{F}_0$. It is clear that the function $\eta^{(k)}(\mu, \boldsymbol{x}_i)$ tends to infinity as $\boldsymbol{x}_i$ approaches a point from boundary of $\mathcal{F}_0$. Now, we prove that each term in (30) is a μ strongly self-concordant function by showing that (23) is satisfied for that term in (30).

First, due to its linearity, it is trivial to show that the linear function $\boldsymbol{c} \bullet \boldsymbol{x}$ is a μ strongly self-concordant barrier on $\mathcal{F}_1$. In fact, (23) is satisfied for any linear function because its first and second Fréchet derivatives are zeros.

Second, we show that the function $-\mu \ln \det \boldsymbol{x}$ is a μ strongly self-concordant barrier on $\mathcal{F}_1$. For $\boldsymbol{\xi} \in \mathcal{S}$, let $\lambda_1\left(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}\right)$ and $\lambda_2\left(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}\right)$ be the eigenvalues of the element $\boldsymbol{x}^{-1} \circ \boldsymbol{\xi} \in \mathcal{S}$. Using (13) and (14), we have

$$\begin{aligned}
\left|\mathcal{D}^3_{\boldsymbol{xxx}}\left(\mu \ln \det \boldsymbol{x}\right)(\boldsymbol{\xi}, \boldsymbol{\xi}, \boldsymbol{\xi})\right| &= 2\mu\ \left|\text{trace}\left(\left(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}\right)^3\right)\right| \\
&= 2\mu\ \left(\left|\lambda_1\left(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}\right)\right|^3 + \left|\lambda_2\left(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}\right)\right|^3\right) \\
&\leq 2\mu\ \left(\lambda_1{}^2\left(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}\right) + \lambda_2{}^2\left(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}\right)\right)^{3/2} \\
&= 2\mu\ \left(\text{trace}\left(\left(\boldsymbol{x}^{-1} \circ \boldsymbol{\xi}\right)^2\right)\right)^{3/2} = 2\mu^{-1/2}\left(\mathcal{D}^2_{\boldsymbol{xx}}\left(\mu \ln \det \boldsymbol{x}\right)(\boldsymbol{\xi}, \boldsymbol{\xi})\right)^{3/2}.
\end{aligned}$$

Thus, (23) is satisfied for the function $\mu \ln \det \boldsymbol{x}$ on $\mathcal{F}_1$.

Third, we show that $\rho^{(k)}(\mu, \boldsymbol{x})$ is a μ strongly self-concordant barrier on $\mathcal{F}_2^{(k)}$ for $k = 1, 2, \cdots, K$. For $\boldsymbol{\xi} \in \mathcal{S}$, we let $\Xi^{(k)}(t) = \mathcal{D}^2_{\boldsymbol{xx}}\rho^{(k)}(\mu, \boldsymbol{x} + t\boldsymbol{\xi})(\boldsymbol{\xi}, \boldsymbol{\xi})$. From Lemma 1, we have

$$|\Xi^{(k)'}(0)| \leq \frac{2}{\sqrt{\mu}}|\Xi^{(k)}(0)|^{3/2}. \tag{31}$$

Note that $\Xi^{(k)}(0) = \mathcal{D}^2_{\boldsymbol{xx}}\rho^{(k)}(\mu, \boldsymbol{x})(\boldsymbol{\xi}, \boldsymbol{\xi})$ and $\Xi^{(k)'}(0) = \mathcal{D}^3_{\boldsymbol{xxx}}\rho^{(k)}(\mu, \boldsymbol{x})(\boldsymbol{\xi}, \boldsymbol{\xi}, \boldsymbol{\xi})$. Therefore, the fact that (31) is satisfied for $\Xi^{(k)}(0)$ on $\mathbb{R}$ implies that (23) is satisfied for $\rho^{(k)}(\mu, \boldsymbol{x})$ on $\mathcal{F}_2^{(k)}$. Thus, $\rho^{(k)}(\mu, \boldsymbol{x})$ is a μ strongly self-concordant barrier on $\mathcal{F}_2^{(k)}$, for $k = 1, 2, \cdots, K$.

Following Proposition 2.1.1(ii) in [25], we conclude that $\sum_{k=1}^{K} \rho^{(k)}(\mu, \boldsymbol{x})$ is a μ strongly self-concordant barrier on $\mathcal{F}_2$. Following the same proposition, the function $\eta^{(k)}(\mu, \boldsymbol{x})$ is a μ strongly self-concordant barrier on $\mathcal{F}_0$. The proof is complete.

Next, we show that the family of functions $\{\eta(\mu, \cdot) : \mu > 0\}$ is a strongly self-concordant family with appropriate parameters. We now introduce the definition of a self-concordant family of functions. Nesterov and Nemirovskii introduced this definition for finite-dimensional optimization (see Definition 3.1.1 in [25]). This definition is introduced below for optimization in a Hilbert space (see also [9]).

Definition 2 *Let C be an open nonempty convex subset of a Hilbert space $\mathcal{H}$. Let also $\mu \in \mathbb{R}_{++}$ and $f_\mu : \mathbb{R}_{++} \otimes C \to \mathbb{R}$ be a family of functions indexed by μ. Let $\alpha_1(\mu)$, $\alpha_2(\mu)$, $\alpha_3(\mu)$, $\alpha_4(\mu)$ and $\alpha_5(\mu) : \mathbb{R}_{++} \to \mathbb{R}_{++}$ be continuously differentiable functions on μ. Then the family of functions $\{f_\mu\}_{\mu\in\mathbb{R}_{++}}$ is called* strongly self-concordant with the parameters α_1, α_2, α_3, α_4, α_5, *if the following conditions hold:*

(i) The function f_μ is continuous on $\mathbb{R}_{++} \otimes C$, and for fixed $\mu \in \mathbb{R}_{++}$, f_μ is convex on C and has three partial Fréchet derivatives on C, which are continuous on $\mathbb{R}_{++} \otimes C$ and continuously differentiable with respect to μ on $\mathbb{R}_{++}$.

(ii) For any $\mu \in \mathbb{R}_{++}$, the function f_μ is strongly $\alpha_1(\mu)$-self-concordant.

(iii) For any $(\mu, \boldsymbol{x}) \in \mathbb{R}_{++} \otimes C$ and any $\boldsymbol{\xi} \in \mathcal{H}$, we have

$$\left|\frac{\partial}{\partial\mu}\left(\mathcal{D}_{\boldsymbol{x}} f_\mu(\mu,\boldsymbol{x})(\boldsymbol{\xi})\right) - \frac{\partial}{\partial\mu}(\ln \alpha_3(\mu))\ \mathcal{D}_{\boldsymbol{x}} f_\mu(\mu,\boldsymbol{x})(\boldsymbol{\xi})\right| \leq \alpha_4(\mu)\,(\alpha_1(\mu))^{1/2}\left(\mathcal{D}^2_{\boldsymbol{xx}} f_\mu(\mu,\boldsymbol{x})(\boldsymbol{\xi},\boldsymbol{\xi})\right)^{1/2}.$$

(iv) For any $(\mu, \boldsymbol{x}) \in \mathbb{R}_{++} \otimes C$ and any $\boldsymbol{\xi} \in \mathcal{H}$, we have

$$\left|\frac{\partial}{\partial\mu}\left(\mathcal{D}^2_{\boldsymbol{xx}} f_\mu(\mu,\boldsymbol{x})(\boldsymbol{\xi},\boldsymbol{\xi})\right) - \frac{\partial}{\partial\mu}(\ln \alpha_2(\mu))\ \mathcal{D}^2_{\boldsymbol{xx}} f_\mu(\mu,\boldsymbol{x})(\boldsymbol{\xi},\boldsymbol{\xi})\right| \leq 2\alpha_5(\mu)\,\mathcal{D}^2_{\boldsymbol{xx}} f_\mu(\mu,\boldsymbol{x})(\boldsymbol{\xi},\boldsymbol{\xi}).$$

The proof of self-concordancy of the family of the recourse functions depends on the following lemma.

Lemma 2 *For any $\mu > 0, \boldsymbol{x} \in \mathcal{F}_0$ and $\boldsymbol{\zeta} \in \mathcal{S}$, the following inequality holds:*

$$\left|\{\mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu,\boldsymbol{x})(\boldsymbol{\zeta},\boldsymbol{\zeta})\}'\right| \leq \frac{\sqrt{r_2}}{\mu}\ \mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu,\boldsymbol{x})(\boldsymbol{\zeta},\boldsymbol{\zeta}). \tag{32}$$

Proof Using (26), (20) and (21), we have

$$\{\mathcal{P}^{(k)^{-2}}(t)\}'\hat{\boldsymbol{e}} = \mathcal{Q}_{\widetilde{\boldsymbol{y}}^{(k)^{-1}}}\left(\hat{\boldsymbol{e}} - 2\,\widetilde{\boldsymbol{s}}^{(k)} \circ \widetilde{\boldsymbol{y}}^{(k)'}\right) = \mathcal{Q}_{\widetilde{\boldsymbol{y}}^{(k)^{-1}}}\left(\hat{\boldsymbol{e}} - 2\,\mathcal{L}(\widetilde{\boldsymbol{s}}^{(k)})\mathcal{P}^{(k)^2}\left(\hat{\jmath} - \widetilde{\mathcal{W}}^{(k)\dagger}R^{(k)^{-1}}\widetilde{\mathcal{W}}^{(k)}\mathcal{P}^{(k)^2}\right)\widetilde{\boldsymbol{y}}^{(k)^{-1}}\right). \tag{33}$$

Note that $\mathcal{P}^{(k)}\widetilde{\mathcal{W}}^{(k)\dagger}R^{(k)^{-1}}\widetilde{\mathcal{W}}^{(k)}\mathcal{P}^{(k)} \succeq 0$. This implies that

$$2\,\mathcal{L}(\widetilde{\boldsymbol{s}}^{(k)})\ \mathcal{P}^{(k)^2}\left(\hat{\jmath} - \widetilde{\mathcal{W}}^{(k)\dagger}R^{(k)^{-1}}\widetilde{\mathcal{W}}^{(k)}\mathcal{P}^{(k)^2}\right)\mathcal{L}\left(\widetilde{\boldsymbol{y}}^{(k)^{-1}}\right) - \hat{\jmath} \preceq \frac{2}{\mu}\ \mathcal{L}(\widetilde{\boldsymbol{s}}^{(k)})\ \mathcal{P}^{(k)^2}\mathcal{L}\left(\widetilde{\boldsymbol{s}}^{(k)}\right) - \hat{\jmath} \preceq \hat{\jmath}. \tag{34}$$

Note also that $\mathcal{Q}_{\widetilde{\boldsymbol{y}}^{(k)}(t)}\left\{\mathcal{P}^{(k)^{-2}}\right\}'$ is a block diagonal matrix and using (33) and (34). It follows for $\boldsymbol{\xi} \in \hat{\mathcal{S}}$ that

$$\left|\boldsymbol{\xi} \bullet \left\{\mathcal{P}^{(k)^{-2}}\right\}'\boldsymbol{\xi}\right| \leq \frac{1}{\mu}\ \left\|\mathcal{Q}_{\widetilde{\boldsymbol{y}}^{(k)}(t)}\left\{\mathcal{P}^{(k)^{-2}}\right\}'\hat{\boldsymbol{e}}\right\|_2 \boldsymbol{\xi} \bullet \mathcal{P}^{(k)^{-2}}\ \boldsymbol{\xi} \leq \frac{1}{\mu}\|\hat{\boldsymbol{e}}\|_F\ \boldsymbol{\xi} \bullet \mathcal{P}^{(k)^{-2}}\ \boldsymbol{\xi} = \frac{\sqrt{r_2}}{\mu}\ \boldsymbol{\xi} \bullet \mathcal{P}^{(k)^{-2}}\ \boldsymbol{\xi}.$$

From (22), it follows for $\boldsymbol{\zeta} \in \mathcal{S}$ that

$$\left|\left\{\mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu,\boldsymbol{x})(\boldsymbol{\zeta},\boldsymbol{\zeta})\right\}'\right| \leq \sqrt{r_2}\ \boldsymbol{\zeta} \bullet \mathcal{Q}_{\boldsymbol{x}^{-1}}\boldsymbol{\zeta} + \frac{\sqrt{r_2}}{\mu}\sum_{k=1}^{K}\boldsymbol{\zeta} \bullet \mathcal{T}^{(k)\dagger}R^{(k)^{-1}}\mathcal{T}^{(k)}\boldsymbol{\zeta} \leq \frac{\sqrt{r_2}}{\mu}\ \mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu,\boldsymbol{x})(\boldsymbol{\zeta},\boldsymbol{\zeta}).$$

Thus, the inequality (32) is satisfied. The proof is complete.

The following important theorem is the counterpart of Theorem 3.2 in [11] and Theorem 2 in [15] for finite-dimensional stochastic second-order cone programming.

Theorem 2 *The family $\{\eta(\mu,\cdot) : \mu > 0\}$ is a strongly self-concordant family with the following parameters*

$$\alpha_1(\mu) = \mu,\ \alpha_2(\mu) = \alpha_3(\mu) = 1,\ \alpha_4(\mu) = \frac{\sqrt{r_1 + Kr_2}}{\mu},\ \alpha_5(\mu) = \frac{\sqrt{r_2}}{2\mu}.$$

Proof It is clear that condition (i) of Definition 2 is satisfied. Theorem 1 shows that condition (ii) is satisfied. To show that condition (iii) is satisfied, we show that the inequality

$$\left|\{\mathcal{D}_x\eta(\mu,\boldsymbol{x})(\boldsymbol{\zeta})\}'\right| \leq \left(\frac{r_1+Kr_2}{\mu}\ \mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu,\boldsymbol{x})(\boldsymbol{\zeta},\boldsymbol{\zeta})\right)^{1/2}$$

holds for any $\mu > 0, \boldsymbol{x} \in \mathcal{F}_0$ and $\boldsymbol{\zeta} \in \mathcal{S}$. For this purpose, we introduce the two multi-block operators $\mathcal{B}, \mathcal{B}^\dagger : \mathcal{S}\otimes\hat{\mathcal{S}}\otimes\cdots\otimes\hat{\mathcal{S}} \longrightarrow \mathcal{S}\otimes\hat{\mathcal{S}}\otimes\cdots\otimes\hat{\mathcal{S}}$ which are defined by

$$\begin{array}{rcl}\mathcal{B}\boldsymbol{\varsigma} &:=& \left(\mathcal{Q}_{\boldsymbol{x}^{-1/2}}\boldsymbol{\zeta}, \mathcal{Q}_{\boldsymbol{y}^{(1)-1/2}}\mathcal{P}^{(1)2}\widetilde{\mathcal{W}}^{(1)\dagger}R^{(1)^{-1}}\mathcal{T}^{(1)}\boldsymbol{\xi}^{(1)},\ldots,\mathcal{Q}_{\boldsymbol{y}^{(K)-1/2}}\mathcal{P}^{(K)2}\widetilde{\mathcal{W}}^{(K)\dagger}R^{(K)^{-1}}\mathcal{T}^{(K)}\boldsymbol{\xi}^{(K)}\right),\\ \mathcal{B}^\dagger\boldsymbol{\varsigma} &:=& \left(\mathcal{Q}_{\boldsymbol{x}^{-1/2}}\boldsymbol{\zeta}, \mathcal{T}^{(1)\dagger}R^{(1)^{-1}}\widetilde{\mathcal{W}}^{(1)}\mathcal{P}^{(1)2}\mathcal{Q}_{\boldsymbol{y}^{(1)-1/2}}\boldsymbol{\xi}^{(1)},\ldots,\mathcal{T}^{(K)\dagger}R^{(K)^{-1}}\widetilde{\mathcal{W}}^{(K)}\mathcal{P}^{(K)2}\mathcal{Q}_{\boldsymbol{y}^{(K)-1/2}}\boldsymbol{\xi}^{(K)}\right),\end{array}$$

for $\boldsymbol{\varsigma} := \left(\boldsymbol{\zeta}, \boldsymbol{\xi}^{(1)},\ldots,\boldsymbol{\xi}^{(K)}\right) \in \mathcal{S}\otimes\hat{\mathcal{S}}\otimes\cdots\otimes\hat{\mathcal{S}}$. Let $\boldsymbol{\varepsilon} := (\boldsymbol{e}, \hat{\boldsymbol{e}},\ldots,\hat{\boldsymbol{e}})$ be the identity of $\mathcal{S}\otimes\hat{\mathcal{S}}\otimes\cdots\otimes\hat{\mathcal{S}}$ ($1+K$ times). Then, from (22), we have

$$\{\mathcal{D}_x\eta(\mu,\boldsymbol{x})\}' = -\boldsymbol{x}^{-1} + \sum_{k=1}^{K}\mathcal{T}^{(k)\dagger}R^{(k)^{-1}}\widetilde{\mathcal{W}}^{(k)}\mathcal{P}^{(k)2}\widetilde{\boldsymbol{y}}^{(k)^{-1}} = \mathcal{B}^\dagger\boldsymbol{\varepsilon}.$$

Note that $\mathcal{P}^{(k)}\mathcal{Q}_{\boldsymbol{y}^{(k)-1}}\mathcal{P}^{(k)} = \frac{1}{\mu}\hat{\mathcal{I}}$. By using this observation and by viewing of (19), we have

$$\begin{array}{rcl}\boldsymbol{\varsigma}\bullet\mathcal{B}^\dagger\mathcal{B}\boldsymbol{\varsigma} &=& \boldsymbol{\zeta}\bullet\mathcal{Q}_{\boldsymbol{x}^{-1}}\boldsymbol{\zeta} + \displaystyle\sum_{k=1}^{K}\boldsymbol{\xi}^{(k)}\bullet\mathcal{T}^{(k)\dagger}R^{(k)^{-1}}\widetilde{\mathcal{W}}^{(k)}\mathcal{P}^{(k)2}\mathcal{Q}_{\boldsymbol{y}^{(k)-1}}\mathcal{P}^{(k)2}\widetilde{\mathcal{W}}^{(k)\dagger}R^{(k)^{-1}}\mathcal{T}^{(k)}\boldsymbol{\xi}^{(k)}\\ &=& \boldsymbol{\zeta}\bullet\mathcal{Q}_{\boldsymbol{x}^{-1}}\boldsymbol{\zeta} + \mu^{-1}\displaystyle\sum_{k=1}^{K}\boldsymbol{\xi}^{(k)}\bullet\mathcal{T}^{(k)\dagger}R^{(k)^{-1}}\mathcal{T}^{(k)}\boldsymbol{\xi}^{(k)} \;=\; \frac{1}{\mu}\,\boldsymbol{\varsigma}\bullet\mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu,\boldsymbol{x})(\boldsymbol{\varsigma}),\end{array}$$

for any $\boldsymbol{\varsigma} = \left(\boldsymbol{\zeta}, \boldsymbol{\xi}^{(1)},\ldots,\boldsymbol{\xi}^{(K)}\right) \in \mathcal{S}\otimes\hat{\mathcal{S}}\otimes\cdots\otimes\hat{\mathcal{S}}$. Since $\boldsymbol{e}\bullet\boldsymbol{e} = \|\boldsymbol{e}\|_F^2$, and likewise $\hat{\boldsymbol{e}}\bullet\hat{\boldsymbol{e}} = \|\hat{\boldsymbol{e}}\|_F^2$, this yields

$$\{\mathcal{D}_x\eta(\mu,\boldsymbol{x})\}'\bullet\left\{\mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu,\boldsymbol{x})\right\}^{-1}\{\mathcal{D}_x\eta(\mu,\boldsymbol{x})\}' = \frac{1}{\mu}\,\boldsymbol{\varepsilon}\bullet\mathcal{B}\{\mathcal{B}^\dagger\mathcal{B}\}^{-1}\mathcal{B}^\dagger\boldsymbol{\varepsilon} \leq \frac{1}{\mu}\,\boldsymbol{\varepsilon}\bullet\boldsymbol{\varepsilon} = \frac{1}{\mu}\,(\boldsymbol{e}\bullet\boldsymbol{e}+K\hat{\boldsymbol{e}}\bullet\hat{\boldsymbol{e}}) = \frac{1}{\mu}\,(r_1+Kr_2). \quad (35)$$

Therefore, we obtain

$$\begin{array}{rcl}\left|\{\mathcal{D}_x\eta(\mu,\boldsymbol{x})\bullet\boldsymbol{\zeta}\}'\right| &\leq& \left(\{\mathcal{D}_x\eta(\mu,\boldsymbol{x})\}'\bullet\left\{\mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu,\boldsymbol{x})\right\}^{-1}\{\mathcal{D}_x\eta(\mu,\boldsymbol{x})\}'\right)^{1/2}\left(\mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu,\boldsymbol{x})(\boldsymbol{\zeta},\boldsymbol{\zeta})\right)^{1/2}\\ &\leq& \left(\dfrac{r_1+Kr_2}{\mu}\ \mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu,\boldsymbol{x})(\boldsymbol{\zeta},\boldsymbol{\zeta})\right)^{1/2}.\end{array}$$

Thus, condition (iii) is satisfied. Lemma 2 shows that condition (iv) is satisfied. The proof is complete.

6 Algorithms

In Section 5, we have shown that the parametric functions $\eta(\mu,\cdot)$ form a strongly self-concordant family with appropriate parameters. This is indeed a very important result because it provides the ability to develop path-following interior-point algorithms for solving our problem. We formally state the proposed algorithm for solving Problem (6) in Algorithm 6.1.

For a given μ, the optimality conditions for the first-stage problem in (10) are

$$\begin{array}{l}\mathcal{D}_x\eta(\mu,\boldsymbol{x}) - \mathcal{A}^\dagger\boldsymbol{\lambda} = \boldsymbol{0}, \mathcal{A}\boldsymbol{x} = \boldsymbol{b},\\ \mathcal{D}_x\eta(\mu,\boldsymbol{x}) \in \mathcal{G}^\perp, \boldsymbol{x} \in \mathcal{G}.\end{array} \quad (36)$$

Let $\boldsymbol{g} := \mathcal{D}_x\eta(\mu,\boldsymbol{x}) - \mathcal{A}^\dagger\boldsymbol{\lambda}$. We find the search directions $\Delta\boldsymbol{x}$ and $\Delta\boldsymbol{\lambda}$, which need to be found in each iteration, by solving the system:

$$\begin{array}{l}\mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu,\boldsymbol{x})\Delta\boldsymbol{x} - \mathcal{A}^\dagger\Delta\boldsymbol{\lambda} = -\boldsymbol{g},\ \mathcal{A}\Delta\boldsymbol{x} = \boldsymbol{0},\\ \mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu,\boldsymbol{x})\ \Delta\boldsymbol{x} + \boldsymbol{g} \in \mathcal{G}^\perp,\ \Delta\boldsymbol{x} \in \mathcal{G}.\end{array} \quad (37)$$

Algorithm 6.1. A logarithmic-barrier algorithm for stochastic conic programming in spin factors

begin algorithm
initialize $\epsilon, \gamma, \theta, \beta, \boldsymbol{x}^0, \mu^0$
ensure: $\epsilon > 0$, $\gamma \in (0,1)$, $\theta > 0$, $\beta > 0$, $\boldsymbol{x}^0 \in \mathcal{F}_0$, $\mu^0 > 0$
  set $\boldsymbol{x} := \boldsymbol{x}^0$, $\mu := \mu^0$
  while $(\mu \geq \epsilon)$
    for $(k = 1; k < K+1; k{+}{+})$
      solve (8) to obtain $\left(\boldsymbol{y}^{(k)}, \boldsymbol{z}^{(k)}, \boldsymbol{s}^{(k)}\right)$
      choose a scaling element $\boldsymbol{p} \in C\left(\boldsymbol{y}^{(k)}, \boldsymbol{s}^{(k)}\right)$
      compute $\left(\widetilde{\boldsymbol{y}}^{(k)}, \widetilde{\boldsymbol{s}}^{(k)}\right)$
    end for
    set $\boldsymbol{g} = \mathcal{D}_{\boldsymbol{x}}\eta(\mu, \boldsymbol{x}) - \mathcal{A}^{\dagger}\boldsymbol{\lambda}$ and compute $\Delta\boldsymbol{x}$ by solving (37)
    compute $\delta(\mu, \boldsymbol{x}) = \sqrt{\frac{1}{\mu}\Delta\boldsymbol{x} \bullet \mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu, \boldsymbol{x})\Delta\boldsymbol{x}}$
    while $(\delta > \beta)$
      set $\boldsymbol{x} := \boldsymbol{x} + \theta\Delta\boldsymbol{x}$
      for $(k = 1; k < K+1; k{+}{+})$
        solve (8) to obtain $\left(\boldsymbol{y}^{(k)}, \boldsymbol{z}^{(k)}, \boldsymbol{s}^{(k)}\right)$
        choose a scaling element $\boldsymbol{p} \in C\left(\boldsymbol{y}^{(k)}, \boldsymbol{s}^{(k)}\right)$
        compute $\left(\widetilde{\boldsymbol{y}}^{(k)}, \widetilde{\boldsymbol{s}}^{(k)}\right)$
      end for
      set $\boldsymbol{g} = \mathcal{D}_{\boldsymbol{x}}\eta(\mu, \boldsymbol{x}) - \mathcal{A}^{\dagger}\boldsymbol{\lambda}$ and compute $\Delta\boldsymbol{x}$ by solving (37)
      compute $\delta(\mu, \boldsymbol{x}) = \sqrt{\frac{1}{\mu}\Delta\boldsymbol{x} \bullet \mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu, \boldsymbol{x})\Delta\boldsymbol{x}}$
    end while
    set $\mu := \gamma\mu$
  end while
  for $(k = 1; k < K+1; k{+}{+})$
    compute $\left(\boldsymbol{y}^{(k)}, \boldsymbol{s}^{(k)}\right)$ by applying inverse scaling to $\left(\widetilde{\boldsymbol{y}}^{(k)}, \widetilde{\boldsymbol{s}}^{(k)}\right)$
  end for
end algorithm

As mentioned in the introduction, the process of solving (37) and finding the search directions $\Delta\boldsymbol{x}$ depends on the given concrete example. In Section 8, we solve (37) in the context of an example in stochastic control.

The algorithm is initialized with a starting point $\boldsymbol{x}^0 \in \mathcal{F}_0$ and a starting value $\mu^0 > 0$ for the barrier parameter μ, and the inner loop of the algorithm is indexed by a factor $\gamma \in (0,1)$. The value of δ measures the proximity of the current point $\boldsymbol{x}$ to the central path, and the value of β is a threshold for that measure. If the current $\boldsymbol{x}$ is too far away from the central path in the sense that $\delta > \beta$, then we solve the Newton system to find a point close to the central path. We then reduce the value of μ by a factor γ and repeat the whole process until the value of μ is within the tolerance ϵ. Consequently, by tracing the central path as μ goes to zero, this procedure will generate a strictly feasible ϵ-optimal solution to (10).

Depending on the input value of γ in Algorithm 6.1, we have two versions of algorithms: The short- and long-step algorithms. More details about these versions of algorithms are given in the next section.

7 Complexity

In this section, we introduce the complexity analysis for the two versions of algorithms. We start with the short-step algorithm.

In the short-step algorithm, a factor $\gamma := 1 - \sigma/\sqrt{r_1 + Kr_2}$, with $\sigma < 0.1$, is used to decrease the barrier parameter μ in each iteration. The k^{th} iteration of the short-step algorithm is executed as follows: At the beginning of the iteration, we have $\mu^{(k-1)}$ and $\boldsymbol{x}^{(k-1)}$ on hand and $\boldsymbol{x}^{(k-1)}$ is close to the central path, i.e., $\delta(\mu^{(k-1)}, \boldsymbol{x}^{(k-1)}) \leq \beta$. After reducing μ from $\mu^{(k-1)}$ to $\mu^k := \gamma\mu^{(k-1)}$, we have $\delta(\mu^k, \boldsymbol{x}^{(k-1)}) \leq 2\beta$. Then a full Newton step with size $\theta = 1$ is taken to produce a new point $\boldsymbol{x}^k$ with $\delta(\mu^k, \boldsymbol{x}^k) \leq \beta$.

The complexity result for the short-step algorithm is given in Theorem 3.

Theorem 3 *Let μ^0 be the initial barrier parameter, $\epsilon > 0$ be the stopping criterion, and $\beta = (2 - \sqrt{3})/2$. If the starting point $\boldsymbol{x}^0$ is sufficiently close to the central path, i.e., $\delta(\mu^0, \boldsymbol{x}^0) \leq \beta$, then the short-step algorithm reduces the barrier parameter μ at a linear rate and terminates with at most $O(\sqrt{r_1 + Kr_2}\ln(\mu^0/\epsilon))$ iterations.*

Proof We determine a good asymptotic upper bound on the number of iterations needed to obtain an ϵ-optimal solution using Algorithm 6.1 when $\gamma = 1 - \sigma/\sqrt{r_1 + Kr_2}$ ($\sigma > 0$). Let N_{out} (respectively, N_{in}) be the number of times we go around the outer (respectively, inner) while loop, and N be the number of iterations needed to an ϵ-optimal solution using Algorithm 6.1. Then, in light of the flowchart in Figure **??**, we have

$$N = N_{\text{in}}\, N_{\text{out}}.$$

We now estimate N_{out}. Let $\mu^{(k)}$ be the parameter at the k^{th} iteration. Then we have

$$\mu^{(k)} = \gamma\mu^{(k-1)} = \gamma^2\mu^{(k-2)} = \cdots = \gamma^k\mu^{(0)}.$$

Thus $\mu^{(k)} < \epsilon$ (here $\epsilon > 0$) if $\gamma^k\mu^{(0)} < \epsilon$. Note that

$$\gamma^k\mu^{(0)} < \epsilon \iff \gamma^k < \epsilon/\mu^{(0)} \iff k\log\gamma = \log(\gamma^k) < \log(\epsilon/\mu^{(0)}).$$

Since $\gamma \in (0, 1)$, we have $k > \log(\epsilon/\mu^{(0)})/\log\gamma = -\log(\mu^{(0)}/\epsilon)/\log\gamma$. It follows that

$$N_{\text{out}} \leq \frac{\log(\mu^{(0)}/\epsilon)}{-\log\gamma}. \tag{38}$$

Since $\gamma = 1 - \sigma/\sqrt{r_1 + Kr_2}$ ($\sigma > 0$), we have[2] $\log\gamma = \log\left(1 - \sigma/\sqrt{r_1 + Kr_2}\right) \approx -\sigma/\sqrt{r_1 + Kr_2}$, and hence

$$N_{\text{out}} \leq \frac{\log(\mu^{(0)}/\epsilon)}{-\log\gamma} \approx \frac{\log(\mu^{(0)}/\epsilon)}{\sigma/\sqrt{r_1 + Kr_2}} = \sqrt{r_1 + Kr_2}\log\left(\frac{\mu^{(0)}}{\epsilon}\right) O(1).$$

From Lemmas 4(i) and 6 given in Appendix B, it follows that we can reduce the parameter μ by the factor $\gamma := 1 - \sigma/\sqrt{r_1 + Kr_2}, \sigma < 0.1$, at each iteration, and that only one Newton step is sufficient to restore proximity to the central path. Therefore $N_{\text{in}} = 1$.

Thus, the number of iterations needed to obtain an ϵ-optimal solution is $N = N_{\text{in}}N_{\text{out}} = O(\sqrt{r_1 + Kr_2}\ln(\mu^0/\epsilon))$. This completes the proof.

In the long-step algorithm, an arbitrary constant factor $\gamma \in (0, 1)$ is used to decrease the barrier parameter μ. It has potential for much faster progress, however, several damped Newton steps might be needed for recentering. The k^{th} iteration of the long-step algorithms is executed as follows: At the beginning of the iteration we have a point $\boldsymbol{x}^{(k-1)}$, which is sufficiently close to $\boldsymbol{x}(\mu^{(k-1)})$, where $\boldsymbol{x}(\mu^{(k-1)})$ is the solution to the first-stage problem in (10) for $\mu := \mu^{(k-1)}$. We reduce the barrier parameter from $\mu^{(k-1)}$ to $\mu^k := \gamma\mu^{(k-1)}$, where $\gamma \in (0, 1)$, and then we search for a point $\boldsymbol{x}^k$ that is sufficiently close to $\boldsymbol{x}(\mu^k)$. In the long-step algorithm, a finite sequence consisting of N points is generated in $\mathcal{F}_0$, and finally $\boldsymbol{x}^k$ is taken to be equal to the last point of this sequence. The complexity result for the long-step algorithm is given in Theorem 4.

[2] For small positive values of x, we have $\log(1 + x) \approx (1 + x) - 1 = x$.

Theorem 4 *Let μ^0 be the initial barrier parameter, $\epsilon > 0$ be the stopping criterion, and $\beta = 1/6$. If the starting point $\boldsymbol{x}^0$ is sufficiently close to the central path, i.e., $\delta(\mu^0, \boldsymbol{x}^0) \leq \beta$, then the long-step algorithm reduces the barrier parameter μ at a linear rate and terminates with at most $\mathcal{O}((r_1 + Kr_2)\ln(\mu^0/\epsilon))$ iterations.*

Proof Let N_{out} (respectively, N_{in}) be the number of times we go around the outer (respectively, inner) while loop. Then the number of iterations needed to obtain an ϵ-optimal solution using Algorithm 6.1 is $N = N_{\text{in}}\, N_{\text{out}}$. Since $\gamma \in (0, 1)$ is an arbitrarily chosen constant, we have $\gamma = \mathcal{O}(1)$. Using (38), it follows that

$$N_{\text{out}} \leq \log\left(\frac{\mu^{(0)}}{\epsilon}\right) \mathcal{O}(1).$$

Now we want to estimate N_{in}. For $\mu > 0$ and $\boldsymbol{x} \in \mathcal{F}_0$, we define $\varrho(\mu, \boldsymbol{x}) := \eta(\mu, \boldsymbol{x}) - \eta(\mu, \boldsymbol{x}(\mu))$. Note that $\varrho(\mu, \boldsymbol{x})$ represents the difference between the minimum objective value $\eta(\mu^k, \boldsymbol{x}(\mu^{k-1}))$ at the beginning of k^{th} iteration and the objective value $\eta(\mu^k, \boldsymbol{x}^{(k)})$ at the end of k^{th} iteration.

Lemma 7 gives upper bounds on $\varrho(\mu, \boldsymbol{x})$ and $\varrho'(\mu, \boldsymbol{x})$. To estimate N_{in}, we first need to find an upper bound on $\varrho(\mu^+, \boldsymbol{x})$. Let $\mu^+ := \gamma\mu$ with $\gamma \in (0, 1)$, and define

$$\tilde{\delta} := \tilde{\delta}(\mu, \boldsymbol{x}) = \sqrt{\frac{1}{\mu}\tilde{\Delta}\boldsymbol{x} \bullet \mathcal{D}^2_{\boldsymbol{xx}}\eta(\mu, \boldsymbol{x})\left(\tilde{\Delta}\boldsymbol{x}\right)}. \tag{39}$$

We particularly show that if $\tilde{\delta} < 1$, then

$$\varrho(\mu^+, \boldsymbol{x}) \leq \mathcal{O}(r_1 + Kr_2)\mu^+. \tag{40}$$

Note that

$$\varrho(\mu, \boldsymbol{x}) = \eta(\mu, \boldsymbol{x}) - \eta(\mu, \boldsymbol{x}(\mu)) = \int_0^1 \mathcal{D}_x\eta(\mu, \boldsymbol{x} + \tau\tilde{\Delta}\boldsymbol{x}) \bullet \tilde{\Delta}\boldsymbol{x}\, d\tau.$$

Since $\boldsymbol{x}(\mu)$ is the optimal solution, we have

$$\mathcal{D}_x\eta(\mu, \boldsymbol{x}(\mu)) = \boldsymbol{0}. \tag{41}$$

Then, for any $\mu > 0$, by applying the chain rule, using (41), and applying the Mean-Value Theorem, we obtain

$$\varrho'(\mu, \boldsymbol{x}) = \eta'(\mu, \boldsymbol{x}) - \eta'(\mu, \boldsymbol{x}(\mu)) - \mathcal{D}_x\eta(\mu, \boldsymbol{x}(\mu)) \bullet \boldsymbol{x}'(\mu) = \mathcal{D}_x\eta\left(\mu, \boldsymbol{x}(\mu) + \varpi\tilde{\Delta}\boldsymbol{x}\right) \bullet \tilde{\Delta}\boldsymbol{x}, \tag{42}$$

Next, by differentiating (42) with respect to μ, we have

$$\varrho''(\mu, \boldsymbol{x}) = \eta''(\mu, \boldsymbol{x}) - \eta''(\mu, \boldsymbol{x}(\mu)) - \mathcal{D}_x\eta'(\mu, \boldsymbol{x}(\mu)) \bullet \boldsymbol{x}'(\mu). \tag{43}$$

We now bound the terms $-\mathcal{D}_x\eta'(\mu, \boldsymbol{x}(\mu)) \bullet \boldsymbol{x}'(\mu)$ and $\eta''(\mu, \boldsymbol{x})$ in (43).

By differentiating $\rho_k(\mu, \boldsymbol{x})$ with respect to μ, using (21) and (9), and using the fact that $\widetilde{\mathcal{W}}^{(k)}\widetilde{\boldsymbol{y}}^{(k)\prime} = \boldsymbol{0}$, we have

$$\rho^{(k)\prime}(\mu, \boldsymbol{x}) = \left(\widetilde{\boldsymbol{d}}^{(k)} - \widetilde{\boldsymbol{s}}^{(k)}\right) \bullet \widetilde{\boldsymbol{y}}^{(k)\prime} - \ln\det\widetilde{\boldsymbol{y}}^{(k)} = \widetilde{\boldsymbol{z}}^{(k)^{\mathsf{T}}}\widetilde{\mathcal{W}}^{(k)}\widetilde{\boldsymbol{y}}^{(k)\prime} - \ln\det\widetilde{\boldsymbol{y}}^{(k)} = -\ln\det\widetilde{\boldsymbol{y}}^{(k)}.$$

Note that $\mathcal{P}^{(k)^2} = \mathcal{P}^{(k)}$. Differentiating $\rho_k'(\mu, \boldsymbol{x})$ with respect to μ and using (21) and (18) give

$$\begin{array}{rcl}
\rho_k''(\mu, \boldsymbol{x}) & = & -\widetilde{\boldsymbol{y}}^{(k)^{-1}} \bullet \widetilde{\boldsymbol{y}}^{(k)\prime} \\
& = & -\hat{\boldsymbol{e}} \bullet \mathcal{L}\left(\widetilde{\boldsymbol{y}}^{(k)^{-1}}\right)\mathcal{P}^{(k)}\left(\hat{\mathcal{I}} - \mathcal{P}^{(k)}\widetilde{\mathcal{W}}^{(k)\dagger}R^{(k)^{-1}}\widetilde{\mathcal{W}}^{(k)}\mathcal{P}^{(k)}\right)\mathcal{P}^{(k)}\mathcal{L}\left(\widetilde{\boldsymbol{y}}^{(k)^{-1}}\right)\hat{\boldsymbol{e}} \\
& \geq & -\|\hat{\boldsymbol{e}}\|_F^2 \left\|\mathcal{P}^{(k)}\mathcal{Q}_{\widetilde{\boldsymbol{y}}^{(k)^{-1}}}\mathcal{P}^{(k)}\right\|_2^2 \;=\; -\dfrac{r_2}{\mu},
\end{array}$$

where the last equality is obtained by noting that $\mathcal{P}^{(k)}\mathcal{Q}_{\widetilde{\boldsymbol{y}}^{(k)^{-1}}}\mathcal{P}^{(k)} = \frac{1}{\mu}\hat{\mathcal{I}}$.

Thus, we have

$$\eta''(\mu, \boldsymbol{x}(\mu)) \geq -\frac{Kr_2}{\mu}. \tag{44}$$

From (36), we have $\mathcal{D}_x\eta(\mu,\boldsymbol{x}) - \mathcal{A}^{\dagger}\boldsymbol{\lambda} = \mathbf{0}$. By differentiating this with respect to μ, we get

$$\mathcal{D}_x\eta'(\mu,\boldsymbol{x}(\mu)) + \mathcal{D}^2_{xx}\eta(\mu,\boldsymbol{x}(\mu))\boldsymbol{x}'(\mu) - \mathcal{A}^{\dagger}\boldsymbol{\lambda}'(\mu) = \mathbf{0}.$$

Consequently, by using (35), we have

$$-\mathcal{D}_x\eta'(\mu,\boldsymbol{x}(\mu)) \bullet \boldsymbol{x}'(\mu) \leq -\mathcal{D}_x\eta'(\mu,\boldsymbol{x}(\mu)) \bullet \left(\mathcal{D}^2_{xx}\eta(\mu,\boldsymbol{x}(\mu))\right)^{-1} \mathcal{D}_x\eta'(\mu,\boldsymbol{x}(\mu)) \leq \frac{1}{\mu}(r_1 + Kr_2). \tag{45}$$

Combining (44) and (45), and using the fact that $\eta''(\mu,\boldsymbol{x}) \leq 0$, we get

$$\varrho''(\mu,\boldsymbol{x}(\mu)) \leq \frac{r_1 + 2Kr_2}{\mu}. \tag{46}$$

By applying the Mean-Value Theorem, and by using Lemma 7, the inequality (46) and the fact that $\gamma^{-1} = \mu^+/\mu \geq \tau/\mu$, we obtain

$$\begin{array}{rcl} \varrho(\mu^+,\boldsymbol{x}) & = & \varrho(\mu,\boldsymbol{x}) + \varrho'(\mu,\boldsymbol{x})(\mu^+ - \mu) + \displaystyle\int_{\mu}^{\mu^+}\int_{\mu}^{\tau} \varrho''(\upsilon,\boldsymbol{x})\, d\upsilon d\tau \\ & \leq & \mu\left(\dfrac{\tilde{\delta}}{1-\tilde{\delta}} + \ln(1-\tilde{\delta})\right) - \sqrt{r_1 + Kr_2}\,\ln(1-\tilde{\delta})\,(\mu - \mu^+) + (r_1 + 2Kr_2)\,(\mu - \mu^+)\,\ln\gamma^{-1}. \end{array}$$

Since $\tilde{\delta}$ and γ are constants, the asymptotic bound in (40) follows.

The aim in what follows is seeking an estimation of N_{in} in light of (40). Assume that $\eta(\mu^{(k-1)},\boldsymbol{x}^{(k-1)}) \leq \beta$ and $\mu^{(k)} = \gamma\mu^{(k-1)}$. Let also

$$\tilde{\boldsymbol{x}}^{(0)} = \boldsymbol{x}^{(k-1)}, \tilde{\boldsymbol{x}}^{(1)} = \boldsymbol{x}^{(k)}, \ldots, \tilde{\boldsymbol{x}}^{(j)} = \boldsymbol{x}^{(j+k-1)}, \ldots$$

be the inner iterates that are generated in the k^{th} loop. We choose β to be small enough so that, in the sense of Lemma 8, $\tilde{\delta} \leq \tilde{\beta} < 1$ at $(\mu^{(k-1)},\boldsymbol{x}^{(k-1)})$.

From (40), there is a positive constant ξ such that

$$\varrho\left(\mu^{(k)},\boldsymbol{x}^{(k-1)}\right) = \eta\left(\mu^{(k)},\boldsymbol{x}^{(k-1)}\right) - \eta\left(\mu^{(k)},\boldsymbol{x}\left(\mu^{(k)}\right)\right) \leq \xi\,(r_1 + Kr_2)\mu^{(k)}.$$

Assume that $\delta^{(i)} := \delta(\mu^{(k)},\tilde{\boldsymbol{x}}^{(i)})$, for all $i = 0,1,\ldots,j-1$. Then the inner loop terminates within j iterations if $\delta^{(j)} \leq \beta$.

Define $\sigma := \beta - \ln(1+\beta) > 0$. It follows that

$$\sigma < \delta^{(i)} - \ln\left(1 + \delta^{(i)}\right), \text{ for all } i = 0,1,\ldots,j-1. \tag{47}$$

Note that

$$\eta\left(\mu^{(k)},\tilde{\boldsymbol{x}}^{(i+1)}\right) \leq \eta\left(\mu^{(k)},\tilde{\boldsymbol{x}}^{(i)} + \theta\Delta\boldsymbol{x}\right) \leq \eta\left(\mu^{(k)},\tilde{\boldsymbol{x}}^{(i)}\right) - \left(\delta^{(i)} - \ln\left(1 + \delta^{(i)}\right)\right)\mu^{(k)} \leq \eta\left(\mu^{(k)},\tilde{\boldsymbol{x}}^{(i)}\right) - \sigma\mu^{(k)},$$

where the second and third inequalities follow from Lemma 4(ii) and (47), respectively. Then we have

$$\eta\left(\mu^{(k)},\boldsymbol{x}\left(\mu^{(k)}\right)\right) \leq \eta\left(\mu^{(k)},\tilde{\boldsymbol{x}}^{(j)}\right) \leq \eta\left(\mu^{(k)},\boldsymbol{x}^{(k-1)}\right) - j\sigma\mu^{(k)} \leq \xi\,(r_1 + Kr_2)\mu^{(k)} - j\sigma\mu^{(k)} + \eta\left(\mu^{(k)},\boldsymbol{x}\left(\mu^{(k)}\right)\right).$$

Therefore $j \leq \xi(r_1+Kr_2)/\sigma$, which means that $\delta^{(j)} \leq \beta$ for any $j > (r_1+Kr_2)/\sigma$. Hence the inner loop terminates within $(r_1 + Kr_2)/\sigma$ iterations. Since σ is a constant, we have $N_{\text{in}} = \mathcal{O}(r_1 + Kr_2)$.

Thus, the number of iterations needed to obtain an ϵ-optimal solution is $N = N_{\text{in}}N_{\text{out}} = \mathcal{O}((r_1 + Kr_2)\ln(\mu^0/\epsilon))$. The result is established.

8 Example

In this section, we present a concrete example from stochastic control theory, and show how this example can be solved with the proposed algorithm.

Let $L_2^n[t_0, t_1]$ be the vector space of measurable square integrable functions on $[t_0, t_1]$ with values in $\mathbb{R}^n$. Define the Hilbert spaces:

$$\mathcal{H} = L_2^{n_1}[t_0, t_1] \otimes L_2^{l_1}[t_0, t_1], \text{ and } \hat{\mathcal{H}} = L_2^{n_2}[t_1, t_2] \otimes L_2^{l_2}[t_1, t_2].$$

In what follows, we consider a two-stage stochastic min-max optimization problem with recourse over the spin factors associated with $\mathcal{H}$ and $\hat{\mathcal{H}}$.

Define the subspaces:

$$\begin{array}{rcl} \mathcal{A} &=& \left\{ \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \in \mathcal{H} : \boldsymbol{\alpha} \text{ is absolutely continuous on } [t_0, t_1], \boldsymbol{\alpha}(t_0) = \mathbf{0}, \dot{\boldsymbol{\alpha}}(t) = A(t)\boldsymbol{\alpha}(t) + B(t)\boldsymbol{\beta}(t), t \in [t_0, t_1] \right\}, \\ \mathcal{W}_\omega &=& \left\{ \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\delta} \end{bmatrix} \in \hat{\mathcal{H}} : \boldsymbol{\gamma} \text{ is absolutely continuous on } [t_1, t_2], \boldsymbol{\gamma}(t_1) = \mathbf{0}, \dot{\boldsymbol{\gamma}}(t) = W_\omega(t)\boldsymbol{\gamma}(t) + T_\omega(t)\boldsymbol{\delta}(t), t \in [t_1, t_2] \right\}, \end{array} \tag{48}$$

where $\dot{\boldsymbol{\alpha}}(t)$ is the time derivative of the function $\boldsymbol{\alpha}(\cdot)$ with respect to time t, the maps $A(t) : \mathbb{R}^{n_1} \longrightarrow \mathbb{R}^{n_1}$, and $B(t) : \mathbb{R}^{l_1} \longrightarrow \mathbb{R}^{n_1}$, $t \in [t_0, t_1]$, are continuous matrix-valued functions, and the maps $W_\omega(t) : \mathbb{R}^{n_2} \longrightarrow \mathbb{R}^{n_2}$, and $T_\omega(t) : \mathbb{R}^{l_2} \longrightarrow \mathbb{R}^{n_2}$, $t \in [t_1, t_2]$, are continuous random matrix-valued functions.

For $i = 1, 2, \ldots, m_1$ and $j = 1, 2, \ldots, m_2$, assume that R_i and S_i are continuous matrix-valued functions on $[t_0, t_1]$, and that $\hat{R}_{\omega,j}$ and $\hat{S}_{\omega,j}$ are continuous random matrix-valued functions on $[t_1, t_2]$ whose realizations depend on an underlying outcome ω in an event space Ω with known probability function P. We deal with a stochastic model of the form:

$$\begin{array}{ll} \min & \displaystyle\max_{1 \le i \le m_1} \int_{t_0}^{t_1} \left((\boldsymbol{\alpha}(t) - \breve{\boldsymbol{\alpha}}_i(t))^{\mathsf{T}} R_i(t) \left(\boldsymbol{\alpha}(t) - \breve{\boldsymbol{\alpha}}_i(t)\right) + \left(\boldsymbol{\beta}(t) - \breve{\boldsymbol{\beta}}_i(t)\right)^{\mathsf{T}} S_i(t) \left(\boldsymbol{\beta}(t) - \breve{\boldsymbol{\beta}}_i(t)\right) \right) dt + \mathbb{E}[Q((\boldsymbol{\alpha}, \boldsymbol{\beta}), \omega)], \\ \text{s.t.} & \displaystyle\sum_{i=1}^{m_1} \left(\int_{t_0}^{t_1} \left(\acute{\boldsymbol{a}}_\iota^{\mathsf{T}}(t) \left(\boldsymbol{\alpha}(t) - \breve{\boldsymbol{\alpha}}_i(t)\right) + \grave{\boldsymbol{a}}_\iota^{\mathsf{T}}(t) \left(\boldsymbol{\beta}(t) - \breve{\boldsymbol{\beta}}_i(t)\right) \right) dt \right) = b_\iota, \ \iota = 1, 2, \ldots, \bar{\iota}, \\ & \begin{bmatrix} \boldsymbol{\alpha} - \breve{\boldsymbol{\alpha}}_i \\ \boldsymbol{\beta} - \breve{\boldsymbol{\beta}}_i \end{bmatrix} \in \mathcal{A}, \ i = 1, 2, \ldots, m_1, \\ & R_i(t) \succeq 0, S_i(t) \succeq 0, \ t \in [t_0, t_1], \ i = 1, 2, \ldots, m_1, \end{array} \tag{49}$$

where $\mathbb{E}[Q((\boldsymbol{\alpha}, \boldsymbol{\beta}), \omega)] = \int_\Omega Q((\boldsymbol{\alpha}, \boldsymbol{\beta}), \omega) P(d\omega)$, and $Q((\boldsymbol{\alpha}, \boldsymbol{\beta}), \omega)$ is the minimum value of the problem

$$\begin{array}{ll} \min & \displaystyle\max_{1 \le j \le m_2} \int_{t_1}^{t_2} \left(\left(\boldsymbol{\gamma}(t) - \breve{\boldsymbol{\gamma}}_{\omega,j}(t)\right)^{\mathsf{T}} \hat{R}_{\omega,j}(t) \left(\boldsymbol{\gamma}(t) - \breve{\boldsymbol{\gamma}}_{\omega,j}(t)\right) + \left(\boldsymbol{\delta}(t) - \breve{\boldsymbol{\delta}}_{\omega,j}(t)\right)^{\mathsf{T}} \hat{S}_{\omega,j}(t) \left(\boldsymbol{\delta}(t) - \breve{\boldsymbol{\delta}}_{\omega,j}(t)\right) \right) dt, \\ \text{s.t.} & \displaystyle\sum_{j=1}^{m_2} \left(\int_{t_1}^{t_2} \left(\acute{\boldsymbol{w}}_{\omega,\varsigma}^{\mathsf{T}}(t) \left(\boldsymbol{\gamma}(t) - \breve{\boldsymbol{\gamma}}_{\omega,j}(t)\right) + \grave{\boldsymbol{w}}_{\omega,\varsigma}^{\mathsf{T}}(t) \left(\boldsymbol{\delta}(t) - \breve{\boldsymbol{\delta}}_{\omega,j}(t)\right) \right) dt \right) = h_{\omega,\varsigma} \\ & \qquad \displaystyle - \sum_{j=1}^{m_2} \left(\int_{t_0}^{t_1} \left(\acute{\boldsymbol{t}}_{\omega,\varsigma}^{\mathsf{T}}(t) \left(\boldsymbol{\alpha}(t) - \breve{\boldsymbol{\alpha}}_i(t)\right) + \grave{\boldsymbol{t}}_{\omega,\varsigma}^{\mathsf{T}}(t) \left(\boldsymbol{\beta}(t) - \breve{\boldsymbol{\beta}}_i(t)\right) \right) dt \right), \ \varsigma = 1, 2, \ldots, \bar{\varsigma}, \\ & \begin{bmatrix} \boldsymbol{\gamma} - \breve{\boldsymbol{\gamma}}_{\omega,j} \\ \boldsymbol{\delta} - \breve{\boldsymbol{\delta}}_{\omega,j} \end{bmatrix} \in \mathcal{W}_\omega, \ j = 1, 2, \ldots, m_2, \\ & \hat{R}_{\omega,j}(t) \succeq 0, \hat{S}_{\omega,j}(t) \succeq 0, \ t \in [t_1, t_2], \ j = 1, 2, \ldots, m_2. \end{array} \tag{50}$$

We obtained the stochastic model (49, 50) by extending a control design model problem (see [8], for example) and applying a relaxation on its assumptions that include deterministic data to include random data. This is a significant problem in stochastic control theory, namely, the problem of stochastic multi-criteria design of the analytic regulator. See Figure 1.

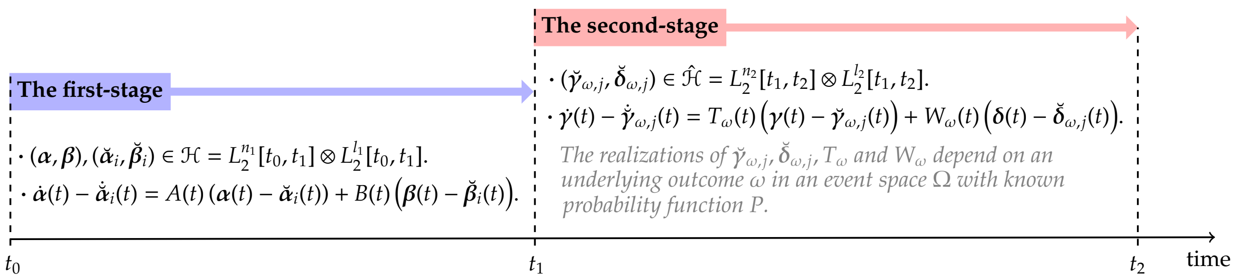

Figure 1: Two-stage stochastic multi-criteria design problem.

The immediate goal is to rewrite Problem (49, 50) in the form of Problem (2). Note that the Hilbert space $\mathcal{H}$ (respectively, $\hat{\mathcal{H}}$) is equipped with the inner product

$$\left\langle \begin{bmatrix} \boldsymbol{\alpha}^{(1)} \\ \boldsymbol{\beta}^{(1)} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\alpha}^{(2)} \\ \boldsymbol{\beta}^{(2)} \end{bmatrix} \right\rangle_{\mathcal{H}\,(\text{respectively}, \hat{\mathcal{H}})} = \int_{t_0\,(\text{resp.}, t_1)}^{t_1\,(\text{resp.}, t_2)} \left(\boldsymbol{\alpha}^{(1)^{\mathsf{T}}}(t)\boldsymbol{\alpha}^{(2)}(t) + \boldsymbol{\beta}^{(1)^{\mathsf{T}}}(t)\boldsymbol{\beta}^{(2)}(t) \right) dt,$$

and is induced by the norm $\|\cdot\|_{\mathcal{H}\,(\text{resp.}, \hat{\mathcal{H}})} = \sqrt{\langle \cdot, \cdot \rangle_{\mathcal{H}\,(\text{resp.}, \hat{\mathcal{H}})}}$.

Using the Cholesky decomposition, we decompose the matrix $R_i(t)$ (respectively, the matrix $S_i(t)$) into a lower triangular matrix $L_{R_i}(t)$ (respectively, $L_{S_i}(t)$) and an upper triangular matrix with transposed values, for $i = 1, 2, \ldots, m_1$. Similarly, we also decompose the matrix $R_{\omega,j}(t)$ (respectively, the matrix $S_{\omega,j}(t)$) into a lower triangular matrix $L_{R_{\omega,j}}(t)$ (respectively, $L_{S_{\omega,j}}(t)$) and an upper triangular matrix with transposed values, for $j = 1, 2, \ldots, m_2$. Then Problem (49, 50) is equivalent to the min-max problem

$$\begin{array}{lll} \min & \max\limits_{1 \le i \le m_1} & \left\| \begin{bmatrix} L_{R_i}^{\mathsf{T}} & 0 \\ 0 & L_{S_i}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} - \breve{\boldsymbol{\alpha}}_i \\ \boldsymbol{\beta} - \breve{\boldsymbol{\beta}}_i \end{bmatrix} \right\|_{\mathcal{H}} + \mathbb{E}[Q((\boldsymbol{\alpha}, \boldsymbol{\beta}), \omega)] \\ \text{s.t.} & & \sum\limits_{i=1}^{m_1} \left\langle \begin{bmatrix} \acute{\boldsymbol{a}}_\iota \\ \grave{\boldsymbol{a}}_\iota \end{bmatrix}, \begin{bmatrix} \boldsymbol{\alpha} - \breve{\boldsymbol{\alpha}}_i \\ \boldsymbol{\beta} - \breve{\boldsymbol{\beta}}_i \end{bmatrix} \right\rangle_{\mathcal{H}} = b_\iota, \ \iota = 1, 2, \ldots, \bar{\iota}, \\ & & \begin{bmatrix} \boldsymbol{\alpha} - \breve{\boldsymbol{\alpha}}_i \\ \boldsymbol{\beta} - \breve{\boldsymbol{\beta}}_i \end{bmatrix} \in \mathcal{A}, \ i = 1, 2, \ldots, m_1, \end{array} \tag{51}$$

where $Q((\boldsymbol{\alpha}, \boldsymbol{\beta}), \omega)$ is the minimum value of the min-max problem

$$\begin{array}{lll} \min & \max\limits_{1 \le j \le m_2} & \left\| \begin{bmatrix} L_{\hat{R}_{\omega,j}}^{\mathsf{T}} & 0 \\ 0 & L_{\hat{S}_{\omega,j}}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma} - \breve{\boldsymbol{\gamma}}_{\omega,j} \\ \boldsymbol{\delta} - \breve{\boldsymbol{\delta}}_{\omega,j} \end{bmatrix} \right\|_{\hat{\mathcal{H}}} \\ \text{s.t.} & & \sum\limits_{j=1}^{m_2} \left\langle \begin{bmatrix} \acute{\boldsymbol{w}}_{\omega,\varsigma} \\ \grave{\boldsymbol{w}}_{\omega,\varsigma} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\gamma} - \breve{\boldsymbol{\gamma}}_{\omega,j} \\ \boldsymbol{\delta} - \breve{\boldsymbol{\delta}}_{\omega,j} \end{bmatrix} \right\rangle_{\hat{\mathcal{H}}} = h_{\omega,\varsigma} - \sum\limits_{j=1}^{m_2} \left\langle \begin{bmatrix} \acute{\boldsymbol{t}}_{\omega,\varsigma} \\ \grave{\boldsymbol{t}}_{\omega,\varsigma} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\alpha} - \breve{\boldsymbol{\alpha}}_i \\ \boldsymbol{\beta} - \breve{\boldsymbol{\beta}}_i \end{bmatrix} \right\rangle_{\mathcal{H}}, \ \varsigma = 1, 2, \ldots, \bar{\varsigma}, \\ & & \begin{bmatrix} \boldsymbol{\gamma} - \breve{\boldsymbol{\gamma}}_{\omega,j} \\ \boldsymbol{\delta} - \breve{\boldsymbol{\delta}}_{\omega,j} \end{bmatrix} \in \mathcal{W}_\omega, \ j = 1, 2, \ldots, m_2. \end{array} \tag{52}$$

We consider the case in which the event space is finite with K realizations. For $k = 1, 2, \ldots, K$, let

$$p_k := P\left(\hat{R}_{\omega,j}, \hat{S}_{\omega,j}, \breve{\boldsymbol{\gamma}}_{\omega,j}, \breve{\boldsymbol{\delta}}_{\omega,j}, \mathcal{W}_\omega, h_\omega, \acute{\boldsymbol{w}}_{\omega,\varsigma}, \grave{\boldsymbol{w}}_{\omega,\varsigma}, \acute{\boldsymbol{t}}_{\omega,\varsigma}, \grave{\boldsymbol{t}}_{\omega,\varsigma}\right) = \left(\hat{R}_{k,j}, \hat{S}_{k,j}, \breve{\boldsymbol{\gamma}}_{k,j}, \breve{\boldsymbol{\delta}}_{k,j}, \mathcal{W}_k, h_k, \acute{\boldsymbol{w}}_{k,\varsigma}, \grave{\boldsymbol{w}}_{k,\varsigma}, \acute{\boldsymbol{t}}_{k,\varsigma}, \grave{\boldsymbol{t}}_{k,\varsigma}\right)$$

be the associated probability. Then Problem (51, 52) with K scenarios can be written in the form

$$\begin{array}{ll}
\min & s + \displaystyle\sum_{k=1}^{K} \bar{\rho}^{(k)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\
\text{s.t.} & \displaystyle\sum_{i=1}^{m_1} \left\langle \begin{bmatrix} \acute{\boldsymbol{a}}_\iota \\ \grave{\boldsymbol{a}}_\iota \end{bmatrix}, \begin{bmatrix} \boldsymbol{\alpha} - \breve{\boldsymbol{\alpha}}_i \\ \boldsymbol{\beta} - \breve{\boldsymbol{\beta}}_i \end{bmatrix} \right\rangle_{\mathcal{H}} = b_\iota,\ \iota = 1, 2, \ldots, \bar{\iota}, \\
& \begin{bmatrix} \boldsymbol{\alpha} - \breve{\boldsymbol{\alpha}}_i \\ \boldsymbol{\beta} - \breve{\boldsymbol{\beta}}_i \end{bmatrix} \in \mathcal{A},\ i = 1, 2, \ldots, m_1, \\
& \left\| \begin{bmatrix} L_{R_i}^{\mathsf{T}} & 0 \\ 0 & L_{S_i}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} - \breve{\boldsymbol{\alpha}}_i \\ \boldsymbol{\beta} - \breve{\boldsymbol{\beta}}_i \end{bmatrix} \right\|_{\mathcal{H}} \leq s,\ i = 1, 2, \ldots, m_1,
\end{array} \tag{53}$$

where $\bar{\rho}^{(k)}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is the minimum value of the problem

$$\begin{array}{ll}
\min & \hat{s} \\
\text{s.t.} & \displaystyle\sum_{j=1}^{m_2} \left\langle \begin{bmatrix} \acute{\boldsymbol{w}}_{k,\varsigma} \\ \grave{\boldsymbol{w}}_{k,\varsigma} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\gamma} - \breve{\boldsymbol{\gamma}}_{k,j} \\ \boldsymbol{\delta} - \breve{\boldsymbol{\delta}}_{k,j} \end{bmatrix} \right\rangle_{\hat{\mathcal{H}}} = h_{k,\varsigma} - \displaystyle\sum_{j=1}^{m_2} \left\langle \begin{bmatrix} \acute{\boldsymbol{t}}_{k,\varsigma} \\ \grave{\boldsymbol{t}}_{k,\varsigma} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\alpha} - \breve{\boldsymbol{\alpha}}_i \\ \boldsymbol{\beta} - \breve{\boldsymbol{\beta}}_i \end{bmatrix} \right\rangle_{\mathcal{H}},\ \varsigma = 1, 2, \ldots, \bar{\varsigma}, \\
& \begin{bmatrix} \boldsymbol{\gamma} - \breve{\boldsymbol{\gamma}}_{k,j} \\ \boldsymbol{\delta} - \breve{\boldsymbol{\delta}}_{k,j} \end{bmatrix} \in \boldsymbol{h}_k + \mathcal{W}_k,\ j = 1, 2, \ldots, m_2, \\
& \left\| \begin{bmatrix} L_{\hat{R}_{k,j}}^{\mathsf{T}} & 0 \\ 0 & L_{\hat{S}_{k,j}}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma} - \breve{\boldsymbol{\gamma}}_{\omega,j} \\ \boldsymbol{\delta} - \breve{\boldsymbol{\delta}}_{\omega,j} \end{bmatrix} \right\|_{\hat{\mathcal{H}}} \leq \hat{s},\ j = 1, 2, \ldots, m_2.
\end{array} \tag{54}$$

Let $\mathcal{S} = \mathbb{R} \otimes \mathcal{H}$ and $\hat{\mathcal{S}} = \mathbb{R} \otimes \hat{\mathcal{H}}$, and define $\Lambda : \mathcal{S} \longrightarrow \mathcal{S} \otimes \mathcal{S} \otimes \cdots \otimes \mathcal{S}$ and $\hat{\Lambda}^{(k)} : \hat{\mathcal{S}} \longrightarrow \hat{\mathcal{S}} \otimes \hat{\mathcal{S}} \otimes \cdots \otimes \hat{\mathcal{S}}$ by

$$\begin{aligned}
\Lambda\left(r, \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}\right) &= \left(\left(r, \begin{bmatrix} L_{R_1}^{\mathsf{T}} & 0 \\ 0 & L_{S_1}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}\right), \left(r, \begin{bmatrix} L_{R_2}^{\mathsf{T}} & 0 \\ 0 & L_{S_2}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}\right), \ldots, \left(r, \begin{bmatrix} L_{R_{m_1}}^{\mathsf{T}} & 0 \\ 0 & L_{S_{m_1}}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}\right)\right), \\
\hat{\Lambda}^{(k)}\left(r, \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\delta} \end{bmatrix}\right) &= \left(\left(r, \begin{bmatrix} L_{\hat{R}_{k,1}}^{\mathsf{T}} & 0 \\ 0 & L_{\hat{S}_{k,1}}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\delta} \end{bmatrix}\right), \left(r, \begin{bmatrix} L_{\hat{R}_{k,2}}^{\mathsf{T}} & 0 \\ 0 & L_{\hat{S}_{k,2}}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\delta} \end{bmatrix}\right), \ldots, \left(r, \begin{bmatrix} L_{\hat{R}_{k,m_2}}^{\mathsf{T}} & 0 \\ 0 & L_{\hat{S}_{k,m_2}}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\delta} \end{bmatrix}\right)\right).
\end{aligned}$$

Let also

$$\begin{array}{lclll}
\boldsymbol{c} & = & ((1, (\mathbf{0}, \mathbf{0})), (0, (\mathbf{0}, \mathbf{0})), \ldots, (0, (\mathbf{0}, \mathbf{0}))) & \in & \mathcal{S} \otimes \mathcal{S} \otimes \cdots \otimes \mathcal{S} \quad (m_1 \text{ times}), \\
\boldsymbol{d} & = & ((1, (\mathbf{0}, \mathbf{0})), (0, (\mathbf{0}, \mathbf{0})), \ldots, (0, (\mathbf{0}, \mathbf{0}))) & \in & \hat{\mathcal{S}} \otimes \hat{\mathcal{S}} \otimes \cdots \otimes \hat{\mathcal{S}} \quad (m_2 \text{ times}), \\
\boldsymbol{a}_\iota & = & ((0, (\acute{\boldsymbol{a}}_\iota, \grave{\boldsymbol{a}}_\iota)), (0, (\acute{\boldsymbol{a}}_\iota, \grave{\boldsymbol{a}}_\iota)), \ldots, (0, (\acute{\boldsymbol{a}}_\iota, \grave{\boldsymbol{a}}_\iota))) & \in & \mathcal{S} \otimes \mathcal{S} \otimes \cdots \otimes \mathcal{S} \quad (m_1 \text{ times}), \\
\boldsymbol{t}_{k,\varsigma} & = & \left(\left(0, \left(\acute{\boldsymbol{t}}_{k,\varsigma}, \grave{\boldsymbol{t}}_{k,\varsigma}\right)\right), \left(0, \left(\acute{\boldsymbol{t}}_{k,\varsigma}, \grave{\boldsymbol{t}}_{k,\varsigma}\right)\right), \ldots, \left(0, \left(\acute{\boldsymbol{t}}_{k,\varsigma}, \grave{\boldsymbol{t}}_{k,\varsigma}\right)\right)\right) & \in & \mathcal{S} \otimes \mathcal{S} \otimes \cdots \otimes \mathcal{S} \quad (m_1 \text{ times}), \\
\boldsymbol{w}_{k,\varsigma} & = & ((0, (\acute{\boldsymbol{w}}_{k,\varsigma}, \grave{\boldsymbol{w}}_{k,\varsigma})), (0, (\acute{\boldsymbol{w}}_{k,\varsigma}, \grave{\boldsymbol{w}}_{k,\varsigma})), \ldots, (0, (\acute{\boldsymbol{w}}_{k,\varsigma}, \grave{\boldsymbol{w}}_{k,\varsigma}))) & \in & \hat{\mathcal{S}} \otimes \hat{\mathcal{S}} \otimes \cdots \otimes \hat{\mathcal{S}} \quad (m_2 \text{ times}).
\end{array}$$

The direct products $\mathcal{S} \otimes \cdots \otimes \mathcal{S}$ and $\hat{\mathcal{S}} \otimes \cdots \otimes \hat{\mathcal{S}}$ are equipped respectively with the inner products

$$\left(\left(r_1^{(1)}, \begin{bmatrix} \boldsymbol{\alpha}_1^{(1)} \\ \boldsymbol{\beta}_1^{(1)} \end{bmatrix}\right), \ldots, \left(r_{m_1}^{(1)}, \begin{bmatrix} \boldsymbol{\alpha}_{m_1}^{(1)} \\ \boldsymbol{\beta}_{m_1}^{(1)} \end{bmatrix}\right)\right) \bullet \left(\left(r_1^{(2)}, \begin{bmatrix} \boldsymbol{\alpha}_1^{(2)} \\ \boldsymbol{\beta}_1^{(2)} \end{bmatrix}\right), \ldots, \left(r_{m_1}^{(2)}, \begin{bmatrix} \boldsymbol{\alpha}_{m_1}^{(2)} \\ \boldsymbol{\beta}_{m_1}^{(2)} \end{bmatrix}\right)\right) = \sum_{i=1}^{m_1} \left(r_i^{(1)} r_i^{(2)} + \left\langle \begin{bmatrix} \boldsymbol{\alpha}_i^{(1)} \\ \boldsymbol{\beta}_i^{(1)} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\alpha}_i^{(2)} \\ \boldsymbol{\beta}_i^{(2)} \end{bmatrix} \right\rangle_{\mathcal{H}}\right),$$

$$\left(\left(r_1^{(1)}, \begin{bmatrix} \boldsymbol{\gamma}_1^{(1)} \\ \boldsymbol{\delta}_1^{(1)} \end{bmatrix}\right), \ldots, \left(r_{m_1}^{(1)}, \begin{bmatrix} \boldsymbol{\gamma}_{m_2}^{(1)} \\ \boldsymbol{\delta}_{m_2}^{(1)} \end{bmatrix}\right)\right) \bullet \left(\left(r_1^{(2)}, \begin{bmatrix} \boldsymbol{\gamma}_1^{(2)} \\ \boldsymbol{\delta}_1^{(2)} \end{bmatrix}\right), \ldots, \left(r_{m_2}^{(2)}, \begin{bmatrix} \boldsymbol{\gamma}_{m_2}^{(2)} \\ \boldsymbol{\delta}_{m_2}^{(2)} \end{bmatrix}\right)\right) = \sum_{j=1}^{m_2} \left(r_j^{(1)} r_j^{(2)} + \left\langle \begin{bmatrix} \boldsymbol{\gamma}_j^{(1)} \\ \boldsymbol{\delta}_j^{(1)} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\gamma}_j^{(2)} \\ \boldsymbol{\delta}_j^{(2)} \end{bmatrix} \right\rangle_{\hat{\mathcal{H}}}\right).$$

Given the above data, we can write Problem (53, 54) in the form

$$\begin{array}{llll}
\min & \boldsymbol{c} \bullet \boldsymbol{x} + \sum_{k=1}^{K} \bar{\rho}^{(k)}(\boldsymbol{x}) & \text{where } \bar{\rho}^{(k)}(\boldsymbol{x}) \text{ is} & \min \quad \boldsymbol{d} \bullet \boldsymbol{y}^{(k)} \\
\text{s.t.} & \mathcal{A}\boldsymbol{x} = \boldsymbol{b}, & \text{the minimum} & \text{s.t.} \quad \mathcal{W}^{(k)} \boldsymbol{y}^{(k)} = \boldsymbol{h}^{(k)} - \mathcal{T}^{(k)} \boldsymbol{x}, \\
& \boldsymbol{x} \in \Lambda(\mathbb{R} \otimes \mathcal{A}), \boldsymbol{x} \geq \mathbf{0}, & \text{value of the problem} & \qquad \boldsymbol{y}^{(k)} \in \hat{\Lambda}^{(k)}\left(\mathbb{R} \otimes \mathcal{W}^{(k)}\right), \boldsymbol{y}^{(k)} \geq \mathbf{0}.
\end{array} \tag{55}$$

Here, based on our notations, $\boldsymbol{x} \geq \mathbf{0}$ ($\boldsymbol{y}^{(k)} \geq \mathbf{0}$) means that $\boldsymbol{x} \in \mathcal{S}^+ \otimes \mathcal{S}^+ \otimes \cdots \otimes \mathcal{S}^+$ ($\boldsymbol{y}^{(k)} \in \hat{\mathcal{S}}^+ \otimes \hat{\mathcal{S}}^+ \otimes \cdots \otimes \hat{\mathcal{S}}^+$), which is equivalent to the second set of constraints in (53) (in (54)).

The next goal now is to write the optimality conditions, which in turns requires writing the dual of the second-stage problem in (55). Note that, from (48), the orthogonal complements of the subspaces $\mathcal{A}$ and

$\mathcal{W}^{(k)}, k = 1, \dots, K$, are

$$\mathcal{A}^{\perp} = \left\{ \begin{bmatrix} \dot{\boldsymbol{p}} + A^{\mathsf{T}}\boldsymbol{p} \\ B^{\mathsf{T}}\boldsymbol{p} \end{bmatrix} : \boldsymbol{p} \text{ is absolutely continuous on } [t_0, t_1], \dot{\boldsymbol{p}}(t) \in L_2^{n_1}[t_0, t_1],\ \boldsymbol{p}(t_1) = \boldsymbol{0} \right\},$$
$$\mathcal{W}^{(k)\perp} = \left\{ \begin{bmatrix} \dot{\boldsymbol{q}}^{(k)} + {W^{(k)}}^{\mathsf{T}}\boldsymbol{q}^{(k)} \\ {T^{(k)}}^{\mathsf{T}}\boldsymbol{q}^{(k)} \end{bmatrix} : \boldsymbol{q}^{(k)} \text{ is absolutely continuous on } [t_1, t_2], \dot{\boldsymbol{q}}^{(k)}(t) \in L_2^{n_2}[t_1, t_2],\ \boldsymbol{q}^{(k)}(t_2) = \boldsymbol{0} \right\}.$$

Consequently, the orthogonal complements of the subspaces $\Lambda(\mathbb{R} \otimes \mathcal{A})$ and $\hat{\Lambda}^{(k)}\left(\mathbb{R} \otimes \mathcal{W}^{(k)}\right), k = 1, \dots, K$, are

$$\Lambda^{\perp}(\mathbb{R} \otimes \mathcal{A}) = \left\{ \left(\left(r_1, \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\beta}_1 \end{bmatrix} \right), \dots, \left(r_{m_1}, \begin{bmatrix} \boldsymbol{\alpha}_{m_1} \\ \boldsymbol{\beta}_{m_1} \end{bmatrix} \right) \right) \in \bigotimes_{i=1}^{m_1} \mathcal{S} : \sum_{i=1}^{m_1} r_i = 0, \sum_{i=1}^{m_1} \begin{bmatrix} L_{R_i}(t) & 0 \\ 0 & L_{S_i}(t) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_i \\ \boldsymbol{\beta}_i \end{bmatrix} \in \mathcal{A}^{\perp} \right\},$$
$$\hat{\Lambda}^{(k)\perp}\left(\mathbb{R} \otimes \mathcal{W}^{(k)}\right) = \left\{ \left(\left(r_1, \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\delta}_1 \end{bmatrix} \right), \dots, \left(r_{m_1}, \begin{bmatrix} \boldsymbol{\gamma}_{m_2} \\ \boldsymbol{\delta}_{m_2} \end{bmatrix} \right) \right) \in \bigotimes_{j=1}^{m_2} \hat{\mathcal{S}} : \sum_{j=1}^{m_2} r_j = 0, \sum_{j=1}^{m_2} \begin{bmatrix} L_{\hat{R}_{k,j}}(t) & 0 \\ 0 & L_{\hat{S}_{k,j}}(t) \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma}_j \\ \boldsymbol{\delta}_j \end{bmatrix} \in \mathcal{W}^{(k)\perp} \right\}.$$

The dual of the second-stage problem in (55) is the problem

$$\begin{array}{ll} \max & \left(\boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x}\right)^{\mathsf{T}} \boldsymbol{z}^{(k)} \\ \text{s.t.} & \boldsymbol{\mathcal{W}}^{(k)\dagger}\boldsymbol{z}^{(k)} + \boldsymbol{s}^{(k)} = \boldsymbol{d}, \\ & \boldsymbol{s}^{(k)} \in \hat{\Lambda}^{(k)\perp}\left(\mathbb{R} \otimes \mathcal{W}^{(k)}\right), \boldsymbol{s}^{(k)} \succeq \boldsymbol{0}. \end{array} \tag{56}$$

It follows that $\boldsymbol{y}^{(k)}$ and $\boldsymbol{s}^{(k)}$ are optimal solutions to the second-stage problem in (55) and Problem (56), respectively, iff they satisfy the following optimality conditions:

$$\begin{array}{l} \boldsymbol{y}^{(k)} \in \hat{\Lambda}^{(k)}\left(\mathbb{R} \otimes \mathcal{W}^{(k)}\right),\ \boldsymbol{s}^{(k)} \in \hat{\Lambda}^{(k)\perp}\left(\mathbb{R} \otimes \mathcal{W}^{(k)}\right), \\ \mathcal{W}^{(k)}\boldsymbol{y}^{(k)} = \boldsymbol{h}^{(k)} - \mathcal{T}^{(k)}\boldsymbol{x},\ \boldsymbol{y}^{(k)} \succ \boldsymbol{0}, \\ \boldsymbol{\mathcal{W}}^{(k)\dagger}\boldsymbol{z}^{(k)} + \boldsymbol{s}^{(k)} = \boldsymbol{d},\ \boldsymbol{s}^{(k)} \succ \boldsymbol{0}. \end{array} \tag{57}$$

Note that the orthogonality conditions in the first line of (57) are equivalent to the equality $\boldsymbol{y}^{(k)} \circ \boldsymbol{s}^{(k)} = \boldsymbol{0}$ (or equivalently, $\boldsymbol{y}^{(k)} \bullet \boldsymbol{s}^{(k)} = 0$). Based on the foregoing development, we can write the barrier problems for Problems (55) and (56), and write the corresponding optimality conditions in the form of (8) and the relaxed optimality conditions in the form of (9). Defining the recourse function $\eta(\mu, \boldsymbol{x})$, we can also compute $\mathcal{D}_x\eta(\mu, \boldsymbol{x})$ and $\mathcal{D}^2_{xx}\eta(\mu, \boldsymbol{x})$, which are calculated in (19).

The major part in computation the search directions is solving System (37). In this case, this system has the form

$$\begin{array}{l} \mathcal{D}^2_{xx}\eta(\mu, \boldsymbol{x})\Delta\boldsymbol{x} - \mathcal{A}^{\dagger}\Delta\boldsymbol{\lambda} = -\boldsymbol{g},\ \mathcal{A}\Delta\boldsymbol{x} = \boldsymbol{0}, \\ \mathcal{D}^2_{xx}\eta(\mu, \boldsymbol{x})\ \Delta\boldsymbol{x} + \boldsymbol{g} \in \Lambda^{\perp}(\mathbb{R} \otimes \mathcal{A}),\ \Delta\boldsymbol{x} \in \Lambda(\mathbb{R} \otimes \mathcal{A}), \end{array} \tag{58}$$

where $\boldsymbol{g} = \mathcal{D}_x\eta(\mu, \boldsymbol{x}) - \mathcal{A}^{\dagger}\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}$ is the first-stage dual multiplier.

The second set of conditions in (58) is equivalent to [8] (see also [10]):

$$\begin{bmatrix} R & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\vartheta} \end{bmatrix} \in \mathcal{A}^{\perp}, \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \in \mathcal{A}, \tag{59}$$

for an appropriate choice of $(\boldsymbol{\theta}, \boldsymbol{\vartheta})$ in $\mathcal{H}$, where $R = \sum_{i=1}^{m_1}\left(x_{i0}^2 - \|\bar{\boldsymbol{x}}_i\|^2\right)R_i$, and $S = \sum_{i=1}^{m_1}\left(x_{i0}^2 - \|\bar{\boldsymbol{x}}_i\|^2\right)S_i$. From the definitions of $\mathcal{A}$ and $\mathcal{A}^{\perp}$, we conclude that System (59) is the system:

$$\begin{array}{rcll} R\boldsymbol{\alpha} + \boldsymbol{\theta} & = & -\dot{\boldsymbol{p}} - A^{\mathsf{T}}\boldsymbol{p}, & \boldsymbol{\alpha}(t_0) = \boldsymbol{0}, \\ S\boldsymbol{\beta} + \boldsymbol{\vartheta} & = & -B^{\mathsf{T}}\boldsymbol{p}, & \boldsymbol{p}(t_1) = \boldsymbol{0}, \\ \dot{\boldsymbol{\alpha}}(0) & = & A\boldsymbol{\alpha} + B\boldsymbol{\beta}. & \end{array} \tag{60}$$

As indicated in [8], System (60) can be solved by solving a matrix differential Riccati equation. To see this, let us try to find $\boldsymbol{p}$ in the form $\boldsymbol{p} = M\boldsymbol{\alpha} + \boldsymbol{\varphi}$, where $M(t_1) = 0, \boldsymbol{\varphi}(t_1) = \boldsymbol{0}$, and $M := M(t)$ is $n_1 \times n_1$ matrix-valued function. Differentiating with respect to t, we get $\dot{\boldsymbol{p}} = \dot{M}\boldsymbol{\alpha} + M\dot{\boldsymbol{\alpha}} + \dot{\boldsymbol{\varphi}}$. By substituting this and the equation in the third line of (60) into the equations in the first two lines in (60), we get

$$\begin{array}{rclrcl} \dot{M} + A^{\mathsf{T}}M + MA - MBS^{-1}B^{\mathsf{T}}M + R & = & 0, & M(t_1) & = & 0, \\ \dot{\boldsymbol{\varphi}} + (A^{\mathsf{T}} - MBS^{-1}B^{\mathsf{T}})\boldsymbol{\varphi} & = & -\boldsymbol{\theta} + MBS^{-1}\boldsymbol{\vartheta}, & \boldsymbol{\vartheta}(t_1) & = & \boldsymbol{0}. \end{array} \tag{61}$$

Interestingly, the first equation in System (61) is a matrix differential Ricatti equation which admits a unique solution in $[t_0, t_1]$ under natural constraints on (A, B). The second equation in System (61) is a finite-dimensional system of linear algebraic equations. The number of linear equations and unknown variables of this system is $m_1 + 1$. The first set of conditions in (58) is also a finite-dimensional system of linear algebraic equations. Therefore, in this case, System (58) can solved efficiently to find the search direction.

9 Conclusions

In this paper, we have presented decomposition-based interior-point algorithms for solving the two-stage stochastic programming problems with recourse over infinite-dimensional second-order cones in spin factors. We have shown that the logarithmic barrier associated with the recourse function is a strongly self-concordant barrier on the first-stage solutions and have accordingly proved the convergence results.

We have also analyzed short- and long-step versions of algorithms that follow the primal central trajectory of the first-stage problem. We have seen that given a direct product of r_1 infinite-dimensional second-order cones in the first-stage problem, a direct product of r_2 infinite-dimensional second-order cones in the second-stage problem, and K number of realizations, we need at most $O((r_1 + Kr_2)^{1/2} \ln(\mu^0/\epsilon))$ Newton iterations in the short-step class of the algorithm to follow the first-stage central path from a starting value of the barrier μ^0 to a terminating value ϵ, and at most $O((r_1 + Kr_2) \ln(\mu^0/\epsilon))$ Newton iterations for this recentering in the long-step class of the algorithm. As an example, we have considered an application of these results to an important stochastic control problem: two-stage stochastic multi-criteria design problem. We have shown that the corresponding infinite-dimensional system can be solved efficiently to find the Newton-type search direction.

In spite of the fact that the result in Theorem 2 is for infinite-dimensional second-order cones and that in [11, Theorem 3.2] is for finite-dimensional second-order cones, we find a remarkable agreement between these two results. This is not surprising, however, because the rank of the JH-algebra associated with the infinite-dimensional second-order cone is finite and is identical to the rank of the Euclidean Jordan algebra associated with the finite-dimensional second-order cone. Note also that the iteration complexity in Theorem 3 for the short-step algorithm and that in Theorem 4 for the long-step algorithm exactly match the iteration complexity of finite-dimensional stochastic second-order cone programming with recourse in [11]. Interestingly, this is also not surprising due to the remarkable agreement that we have found between Theorem 2 in this paper and Theorem 3.2 in [11].

After this work, it has become clear how to generalize other stochastic interior-point algorithms analyzed earlier in [13, 15, 16] for finite-dimensional stochastic second-order cone programming to the infinite-dimensional case in spin factors. We believe that proposed algorithm has performance benefits over any other algorithms for stochastic convex programs in infinite dimensional spaces, and it is attractive from the decompositional and analytical points of view. This attractiveness is due to the exploitation of the special algebraic structure of the infinite-dimensional second-order cone which allowed us to explicitly compute expressions for the derivatives of the barriers and identify the barrier parameters for the self-concordant family. However, in spite of its attractiveness, the algorithm has some limitations which include, but not limited to: producing a good starting point, developing a practical step length selection procedure, reducing μ with a practical strategy in our setting, and selecting a proper choice of ϵ that terminates the algorithm. These limitations can be addressed in a future research study of practical implementations.

References

1. B.S. Mordukhovich, and T.T. Nghia, Second-order variational analysis and characterizations of tilt-stable optimal solutions in infinite-dimensional spaces, Nonlinear Anal., 86 (2013), pp. 159–180.

2. P. Kogut, and R. Manzo, On quadratic scalarization of vector optimization problems in Banach spaces, Appl. Anal., 93 (2014), pp. 994–1009.

3. T.Q. Son, and C.F. Wen, Weak-subdifferentials for vector functions and applications to multiobjective semi-infinite optimization problems, Appl. Anal., 99 (2020), pp. 840–855.

4. C. Qiu, X. Yang, and Y. Zhou, Solvable optimization problems involving a p-Laplacian type operator, Appl. Anal., (2020), DOI: 10.1080/00036811.2020.1842372

5. Yu Han, A Hausdorff-type distance, the Clarke generalized directional derivative and applications in set optimization problems, Appl. Anal., (2020), DOI: 10.1080/00036811.2020.1778673

6. M. Feng, and S. Li, On second-order Fritz John type optimality conditions for a class of differentiable optimization problems, Appl. Anal., 99 (2020), pp. 2594–2608.

7. A.B. Lim, and J.B. Moore, A path following algorithm for infinite quadratic programming on a Hilbert space, Discrete Cont. Dyn. Syst., 4 (1998), pp. 653–670.

8. L. Faybusovich, and T. Tsuchiya, Primal-dual algorithms and infinite-dimensional Jordan algebras of finite rank, Math. Program. Ser. B, 97 (2003), pp. 471–493.

9. J. Renegar, Linear programming, complexity theory and elementary functional analysis, Math. Program., 70 (1995), pp. 279–351.

10. L. Faybusovich, and J.B. Moore, Infinite-dimensional quadratic optimization: Interior-point methods and control applications, Appl. Math. Optim., 36 (1997), pp. 43–66.

11. B. Alzalg, Decomposition-based interior-point methods for stochastic quadratic second-order cone programming, Appl. Math. Comput., 249 (2014), pp. 1–18.

12. C-H. Chu, Infinite dimensional Jordan algebras and symmetric cones, J. Algebra, 491 (2017), pp. 357–371.

13. B. Alzalg, Homogeneous self-dual algorithms for stochastic second-order cone programming, J. Optim. Theory Appl., 163 (2014), pp. 148–164.

14. B. Alzalg, and K.A. Ariyawansa, Logarithmic barrier decomposition-based interior-point methods for stochastic symmetric programming, J. Math. Anal. Appl., 409 (2014), pp. 973–995.

15. B. Alzalg, Volumetric barrier decomposition algorithms for stochastic quadratic second-order cone programming, Appl. Math. Comput., 256 (2015), pp. 494–508.

16. B. Alzalg, K. Badarneh, and A. Ababneh, An infeasible interior-point algorithm for stochastic second-order cone optimization, J. Optim. Theory Appl., 163 (2018), pp. 148–164.

17. B. Alzalg, Logarithmic-barrier decomposition interior-point methods for stochastic linear optimization in a Hilbert space, Numer. Funct. Anal. Optim., 41 (2020), pp. 901–928.

18. C-H. Chu, Jordan Structures in Geometry and Analysis, Cambridge Tracts in Math., vol.190, Cambridge Univ. Press, Cambridge, 2012.

19. T. Nomura, Grassmann manifold of a JH-algebra, Ann. Global Anal. Geom., 12 (1994), pp. 237–260.

20. S.H. Schmieta, and F. Alizadeh, Extension of primal-dual interior point methods to symmetric cones, Math. Program. Ser. A, 96 (2003), pp. 409–438.

21. F. Alizadeh, and D. Goldfarb, Second-order cone programming, Math. Program. Ser. B, 95 (2003), pp. 3–51.

22. C. Helmberg, F. Rendl, R.J. Vanderbei, and H. Wolkowicz, An interior-point methods for stochastic semidefinite programming, SIAM J. Optim., 6 (1996), pp. 342–361.

23. R. Monteiro, Primal-dual path-following algorithms for semidefinite programming, SIAM J. Optim., 7 (1997), pp. 663–678.

24. M. Kojima, S. Shindoh, and S. Hara, Interior-point methods for the monotone linear complementarity problem in symmetric matrices, SIAM J. Optim., 7 (1997), pp. 86–125.

25. Yu.E. Nesterov, and A.S. Nemirovskii, Interior Point Polynomial Algorithms in Convex Programming, SIAM Publications, Philadelphia, PA, 1994.

26. H. Upmeier, Symmetric Banach Manifolds and Jordan $C^\star$-Algebras, North-Holl. Math. Stud., vol. 104, North Holland, Amsterdam, 1985.

27. W. Kaup, Jordan algebras and holomorphy. Functional analysis, holomorphy, and approximation theory, (Proc. Sem., Univ. Fed. Rio de Janeiro, Rio de Janeiro, 1978), 341–365, Lecture Notes in Mathematics, Vol. 843, Springer, Berlin, 1981.

A Basics of Jordan-Hilbert algebras

This appendix complements and expands the material introduced in Section 2. In this appendix, we review some basic definitions and notions of Jordan algebra and Hilbert space, and present some properties that were used in the paper. These properties are associated with the JH-algebra associated with the infinite dimensional second-order cone. Those who are not familiar with the basics of the theory of this algebra are invited to read [18, 26, 27] for more details.

As mentioned earlier in the introduction, Chu [12] has proved that there is a one-one correspondence between unital JH-algebras and infinite-dimensional symmetric cones. To state and utilize this result in our framework, we first introduce the definitions of the unital JH-algebras, their cone of squares, and the infinite-dimensional symmetric cones.

The vector space $\mathcal{J}$ is called an *algebra* over $\mathbb{R}$ if a bilinear map $\diamond : \mathcal{J} \times \mathcal{J} \longrightarrow \mathcal{J}$ exists. An algebra $(\mathcal{J}, \diamond)$ is called a *real Jordan algebra* if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{J}$ we have $\boldsymbol{x} \diamond \boldsymbol{y} = \boldsymbol{y} \diamond \boldsymbol{x}$ (commutativity) and $\boldsymbol{x} \diamond ((\boldsymbol{x} \diamond \boldsymbol{x}) \diamond \boldsymbol{y}) = (\boldsymbol{x} \diamond \boldsymbol{x}) \diamond (\boldsymbol{x} \diamond \boldsymbol{y})$ (Jordan's axiom).

Definition 3 (cf. [19, Definition 1.1]) *Let $\mathcal{J}$ be a real vector space which can be infinite-dimensional. Then:*

1. *A Jordan algebra $(\mathcal{J}, \diamond)$ is called a* real Jordan-Hilbert algebra *(*JH-algebra *for short) if it is a Hilbert space in which its inner product, say "$\blacklozenge$", is associative, i.e., for any $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{J}$, we have $(\boldsymbol{x} \diamond \boldsymbol{y}) \blacklozenge \boldsymbol{z} = \boldsymbol{y} \blacklozenge (\boldsymbol{x} \diamond \boldsymbol{z})$.*
2. *A JH-algebra is called* unital *if it contains an identity, say $\boldsymbol{e}$, which is the unique element in $(\mathcal{J}, \diamond)$ satisfying $\boldsymbol{x} \diamond \boldsymbol{e} = \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathcal{J}$.*
3. *The* cone of squares *of a unital JH-algebra $(\mathcal{J}, \diamond, \blacklozenge)$ is defined as $\mathcal{J}^+ := \{\boldsymbol{x} \diamond \boldsymbol{x} : \boldsymbol{x} \in \mathcal{J}\}$.*

A cone $\mathcal{C}$ is called *regular* if $\mathcal{C}$ is a pointed closed convex cone with non-empty interior.

Definition 4 (cf. [12, Definition 2.4]) *Let $\mathcal{H}$ be a real Hilbert space, which can be infinite-dimensional, with an inner product $\langle \cdot, \cdot \rangle$. Let also $\mathcal{C}$ be a regular cone subset of $\mathcal{H}$. Then:*

1. *The cone $\mathcal{C}$ is called* self-dual *if $\mathcal{C} = \mathcal{C}^\star$, where $\mathcal{C}^\star$ is the* dual cone *of the cone $\mathcal{C}$, and is defined as $\mathcal{C}^\star := \{\boldsymbol{\zeta} \in \mathcal{H} : \langle \boldsymbol{\zeta}, \boldsymbol{x} \rangle \geq 0, \forall \boldsymbol{x} \in \mathcal{C}\}$.*
2. *The cone $\mathcal{C}$ is called* homogeneous *if for any $\boldsymbol{x}, \boldsymbol{y} \in int\, \mathcal{C}$, there is a linear continuous automorphism $g : \mathcal{H} \rightarrow \mathcal{H}$ such that $g(\boldsymbol{x}) = \boldsymbol{y}$.*
3. *The cone $\mathcal{C}$ is called* symmetric *if it is both self-dual and homogeneous.*

The following theorem generalizes to the infinite-dimensional setting the well-known result (cf. [20, Theorem 2]) that assets we have a one-one correspondence between Euclidean Jordan algebras and finite-dimensional symmetric cones.

Theorem 5 (Chu [12, Theorem 3.1]) *A (possibly infinite-dimensional) regular cone is symmetric if and only if it is the cone of squares of a (necessarily unique) unital JH-algebra.*

It is clear that the infinite-dimensional second order cone $\mathcal{K}$ is regular. So, one way to prove that $\mathcal{K}$ is symmetric is to show it is the cone of squares of some unital JH-algebra.

Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a real Hilbert space, which can be infinite-dimensional, with an inner product $\langle\cdot,\cdot\rangle$, identity operator $\mathcal{I}_{\mathcal{H}}$, and zero element $\mathbf{0}$. Let also $\mathcal{S} = \mathcal{S}_{\mathcal{H}} = \mathbb{R} \otimes \mathcal{H}$ be the corresponding spin factor, and $\mathcal{K} = \mathcal{K}_{\mathcal{H}}$ be the corresponding second-order cone defined in (1). For $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$, the map $\circ : \mathcal{S} \otimes \mathcal{S} \longrightarrow \mathcal{S}$ is defined as $\boldsymbol{x} \circ \boldsymbol{y} = (\boldsymbol{x} \bullet \boldsymbol{y}, x_0\bar{y} + y_0\bar{x})$, where the inner product "$\bullet$", which was defined in (3), can be also redefined as $\boldsymbol{x} \bullet \boldsymbol{y} := \frac{1}{2}\text{trace}(\boldsymbol{x} \circ \boldsymbol{y})$.

Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{S}$, and $\alpha, \beta \in \mathbb{R}$. It is not hard to see that $\boldsymbol{x} \circ (\alpha \boldsymbol{y} + \beta \boldsymbol{z}) = \alpha(\boldsymbol{x} \circ \boldsymbol{y}) + \beta(\boldsymbol{y} \circ \boldsymbol{z})$, and $(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) \circ \boldsymbol{z} = \alpha(\boldsymbol{x} \circ \boldsymbol{z}) + \beta(\boldsymbol{y} \circ \boldsymbol{z})$. This means that the map "$\circ$" is bilinear, and hence the structure $(\mathcal{S}, \circ)$ is an algebra. One can also prove that $\boldsymbol{x} \circ \boldsymbol{y} = \boldsymbol{y} \circ \boldsymbol{x}$, and $\boldsymbol{x} \circ (\boldsymbol{x}^{(2)} \circ \boldsymbol{y}) = \boldsymbol{x}^{(2)} \circ (\boldsymbol{x} \circ \boldsymbol{y})$, where $\boldsymbol{x}^{(2)}$ is defined as $\boldsymbol{x}^{(2)} := \boldsymbol{x} \circ \boldsymbol{x}$. This implies that $(\mathcal{S}, \circ)$ is a Jordan algebra. Further, one can also prove that $(\boldsymbol{x} \circ \boldsymbol{y}) \bullet \boldsymbol{z} = \boldsymbol{y} \bullet (\boldsymbol{x} \circ \boldsymbol{z})$, for any $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{S}$, which in turn implies that $(\mathcal{S}, \circ, \bullet)$ is a JH-algebra.

The element $\boldsymbol{e} = (1, \mathbf{0})$ is the identity of the algebra $(\mathcal{S}, \circ)$. The above discussion concludes the following theorem.

Theorem 6 *The algebra $(\mathcal{S}, \circ, \bullet)$ is a unital JH-algebra.*

Although many of the notions and results in this appendix may hold true for general infinite-dimensional Jordan algebras of finite rank, in the subsequent development we deal with the unital JH-algebra $(\mathcal{S}, \circ, \bullet)$ exclusively, because this generality is not needed for this paper.

The following theorem gives the JH-algebraic characterization of the cone $\mathcal{K}$ and states that the infinite-dimensional second-order cone $\mathcal{K}$ defined in (1) is nothing but the cone of squares $\mathcal{S}^+$. We can prove this theorem by following the same arguments as the proof of the fact that the (finite-dimensional) second-order cone $\mathcal{Q}$ (defined in the introduction) is the cone of squares of the Euclidean Jordan algebra $\mathbb{R} \times \mathbb{R}^{n-1}$ (see for example [21, Section 4]). So the proof was omitted, avoiding an unnecessary lengthening of the paper.

Theorem 7 $\mathcal{K} = \mathcal{S}^+$.

Faybusovich and Tsuchiya presented in [8, Propositions 15] the self-duality part of the following corollary by providing a direct proof using the definition. They have also presented in [8, Propositions 7] the homogeneity part of the corollary without proof. This corollary arrives at the same results but from a different angle. The result in the following corollary follows immediately from Theorems 5, 6 and 7.

Corollary 1 *The infinite-dimensional second order cone $\mathcal{K}$ is symmetric, and hence it is a self-dual and homogeneous cone.*

Let $\boldsymbol{x} \in \mathcal{S}$. The decomposition $x = \lambda_1(\boldsymbol{x})\boldsymbol{c}_1(\boldsymbol{x}) + \lambda_2(\boldsymbol{x})\boldsymbol{c}_2(\boldsymbol{x})$, where $\lambda_{1,2}(\boldsymbol{x}) = x_0 \pm \|\bar{x}\|$ and $\boldsymbol{c}_{1,2}(\boldsymbol{x}) = \frac{1}{2}(1, \pm\bar{x}/\,\|\bar{x}\|)$, is the spectral decomposition associated with each element $\boldsymbol{x}$. We can easily see that

$$\boldsymbol{c}_1(\boldsymbol{x}) + \boldsymbol{c}_2(\boldsymbol{x}) = \boldsymbol{e},\ \boldsymbol{c}_1(\boldsymbol{x}) \bullet \boldsymbol{c}_2(\boldsymbol{x}) = 0,\ \boldsymbol{c}_1(\boldsymbol{x}) \circ \boldsymbol{c}_2(\boldsymbol{x}) = \mathbf{0},\ \boldsymbol{c}_1^2(\boldsymbol{x}) = \boldsymbol{c}_1(\boldsymbol{x}),\ \text{and}\ \boldsymbol{c}_2^2(\boldsymbol{x}) = \boldsymbol{c}_2(\boldsymbol{x}), \tag{62}$$

which means that the pair $\{\boldsymbol{c}_1(\boldsymbol{x}), \boldsymbol{c}_2(\boldsymbol{x})\}$ is a Jordan frame.

The elements $\boldsymbol{x}$ and $\boldsymbol{y}$ are *simultaneously decomposed* if they share a Jordan frame, i.e., $\boldsymbol{x} = \lambda_1\boldsymbol{c}_1 + \lambda_2\boldsymbol{c}_2$ and $\boldsymbol{y} = \omega_1\boldsymbol{c}_1 + \omega_2\boldsymbol{c}_2$ for a Jordan frame $\{\boldsymbol{c}_1, \boldsymbol{c}_2\}$. We say $\boldsymbol{x}$ and $\boldsymbol{y}$ *operator commute* if $\boldsymbol{x} \circ (\boldsymbol{y} \circ \boldsymbol{z}) = \boldsymbol{y} \circ (\boldsymbol{x} \circ \boldsymbol{z})$ for all $\boldsymbol{z} \in \mathcal{S}$.

The discussion given in [20, Subsection 2.2] for finite-dimensional Jordan algebras can be extended word-by-word to infinite-dimensional Jordan algebras with finite ranks such as the unital JH-algebra $(\mathcal{S}, \circ, \bullet)$ introduced above. In particular, we can prove Theorem 8 below by following the same arguments as the proof of Theorem 27 in [20]. So, we have decided to omit the proof of this theorem as well, avoiding an unnecessary lengthening of the paper.

Theorem 8 *Two elements of $\mathcal{S}$ operator commute iff they are simultaneously decomposed.*

Two elements $\boldsymbol{x} = (\boldsymbol{x}_1; \boldsymbol{x}_2; \ldots; \boldsymbol{x}_r)$ and $\boldsymbol{y} = (\boldsymbol{y}_1; \boldsymbol{y}_2; \ldots; \boldsymbol{y}_r)$ operator commute iff $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ operator commute for all $i = 1, 2, \ldots, r$.

The image of $\boldsymbol{x} \in \mathcal{S}$ under a smooth function f is $f(\boldsymbol{x}) = f(\lambda_1(\boldsymbol{x}))\,\boldsymbol{c}_1(\boldsymbol{x}) + f(\lambda_2(\boldsymbol{x}))\,\boldsymbol{c}_2(\boldsymbol{x})$. In particular, the *square* of $\boldsymbol{x}$ is given by $\boldsymbol{x}^2 := \lambda_1^2(\boldsymbol{x})\boldsymbol{c}_1(\boldsymbol{x}) + \lambda_2^2(\boldsymbol{x})\boldsymbol{c}_2(\boldsymbol{x}) = (\boldsymbol{x} \bullet \boldsymbol{x}, 2x_0\bar{x}) = \boldsymbol{x}^{(2)}$. In general, for any nonnegative integer k, one can show that $x^k = x^{(k)}$. Using (62), we have for any integers $p, q \geq 1$, $x^p \circ x^q = x^{p+q}$. This means that $(\mathcal{S}, \circ)$ is power associative.

Below are some properties of the operators $\mathcal{L}(\cdot)$, $\mathcal{Q}.$ and $Q_{\cdot,\cdot}$ that we introduced in Section 2. The following lemma is needed to write the complementarity conditions in an equivalent way. It can be proved in the same way as Lemma 28 in [20] and therefore the proof is omitted.

Lemma 3 *Let $x, y \in \mathcal{S}^+$ and p be an invertible element in $\mathcal{S}$. Then $x \circ y = e$ iff $\mathcal{Q}_p x \circ \mathcal{Q}_{p^{-1}} y = e$.*

A derivation in $\mathcal{S}$ is a linear transformation such that if $x, y \in \mathcal{S}$ are functions of $t \in \mathbb{R}$, we have $(x \circ y)' = x \circ y' + x' \circ y$. In other words, $(\mathcal{L}(x)y)' = \mathcal{L}(x)y' + \mathcal{L}(x')y$. Therefore, $(\mathcal{L}(x))' = \mathcal{L}(x')$, $(Q_{x,y})' = Q_{x,y'} + Q_{x',y}$, and, in particular, $(Q_x)' = Q_{x,x'} + Q_{x',x}$.

B Technical lemmas

In this appendix, we give some technical lemmas that were used to estimate N_{in} in Section 7. The following lemma describes the behavior of the Newton method as applied to $\eta(\mu, \cdot)$. It was used in the proofs of both Theorem 3 and Theorem 4.

Lemma 4 *For any $\mu > 0, x \in \mathcal{F}_0$, and Δx, let $\delta := \delta(\mu, x)$ be defined as in line* (10) *of Algorithm 6.1. Then the following relations hold:*

(i) If $\delta < 2 - \sqrt{3}$, then $\delta(\mu, x + \Delta x) \leq (\delta/(1-\delta))^2 \leq \delta/2$.

(ii) If $\delta \geq 2 - \sqrt{3}$, then $\eta(\mu, x) - \eta(\mu, x + \bar{\theta}\Delta x) \geq \mu(\delta - \ln(1+\delta))$, where $\bar{\theta} = (1+\delta)^{-1}$.

The proof of the first part of Lemma 4 follows exactly the proof Proposition 3.4 in [10], and the proof of the second part follows exactly that of Proposition 3.3 in [10]. Note that, in the finite-dimensional case, Lemma 4 is essentially Theorem 2.2.3 in [25].

The result in the following lemma is a restatement of Theorem 3.4 in [10] for our setting. It will be used in the proof of Lemma 6.

Lemma 5 *Let*

$$\chi_\kappa(\eta; \mu, \mu^+) := \left(\frac{1 + r_2}{2} + \frac{\sqrt{r_1 + K r_2}}{\kappa} \right) \ln \gamma^{-1}.$$

Assume that $\delta(\mu, x) < \kappa$ and $\mu^+ := \gamma\mu$ satisfies $\chi_\kappa(\eta; \mu, \mu^+) \leq 1 - \delta(\mu, x)/\kappa$. Then $\delta(\mu^+, x) < \kappa$.

The following lemma was used in the proof of Theorem 3.

Lemma 6 *Let $\mu^+ = \gamma\mu$, where $\gamma = 1 - \sigma/\sqrt{r_1 + K r_2}$ and $\sigma \leq 0.1$, and let $\beta = (2 - \sqrt{3})/2$. If $\delta(\mu, x) \leq \beta$, then $\delta(\mu^+, x) \leq 2\beta$.*

Proof Let $\kappa := 2\beta = 2 - \sqrt{3}$. Since $\delta(\mu, x) \leq \kappa/2$, one can verify that for $\sigma \leq 0.1$, μ^+ satisfies $\chi_\kappa(\eta; \mu, \mu^+) \leq 1/2 \leq 1 - \delta(\mu, x)/\kappa$. By Lemma 5, we have $\delta(\mu^+, x) \leq \kappa$.

The proof of the following lemma follows exactly that of Lemma 10 in [14]. We only indicate that this proof makes use of the result in Theorem 1. Lemma 7 was used in the proof of Theorem 4.

Lemma 7 *Let $\mu > 0$ and $x \in \mathcal{F}_0$, we denote $\tilde{\Delta} x := x - x(\mu)$. For any $\mu > 0$ and $x \in \mathcal{F}_0$, if $\tilde{\delta} < 1$, where $\tilde{\delta}$ is computed in* (39)*, then the following inequalities hold:*

$$\begin{aligned} \varrho(\mu, x) &\leq \mu(\tilde{\delta}/(1-\tilde{\delta}) + \ln(1-\tilde{\delta})), \\ |\varrho'(\mu, x)| &\leq -\sqrt{r_1 + K r_2} \ln(1-\tilde{\delta}). \end{aligned}$$

The following lemma was used in the proof of Theorem 4.

Lemma 8 *For any $\mu > 0, x \in \mathcal{F}_0$, and $(\Delta x, \Delta\lambda)$, let $(\tilde{\Delta} x, \tilde{\Delta}\lambda) := (x - x(\mu), \lambda - \lambda(\mu))$. Let also $\delta = \delta(\mu, x)$ be computed from line (10) of Algorithm 6.1, and $\tilde{\delta} = \tilde{\delta}(\mu, x)$ be computed from* (39)*. If $\delta < 1/6$, then $\frac{2}{3}\delta \leq \tilde{\delta} \leq 2\delta$.*

The proof of Lemma 8 follows that of Lemma 9 in [14].