An inexact proximal bundle method with applications to convex conic programming

CHEK BENG CHUA[*] AND HUILING LIN[†]

ABSTRACT. We present an inexact bundle method for minimizing an unconstrained convex sup-function with an open domain. Under some mild assumptions, we reformulate a convex conic programming problem as such problem in terms of the support function. This method is a first-order method, hence it requires much less computational cost in each iteration than second-order approaches such as interior-point methods. While sometimes providing solutions of low accuracy, such method can attack large scale problems. We show the global convergence of this method. Finally, we give an explicit model for the objective function and a concrete routine to compute the largest eigenvalue inexactly in the case of convex quadratic symmetric cone programming.

1. INTRODUCTION

Let $\mathbb{E}$ denote a Euclidean space with inner product $\langle \cdot, \cdot \rangle$. In this paper, we consider the unconstrained minimization

$$\text{(Min)} \qquad \inf_{x \in \mathbb{E}} f(x),$$

where the domain $dom f$ is an open convex subset of $\mathbb{E}$, and f is the convex sup-function

$$x \in \mathbb{E} \mapsto \sup_{w \in \mathcal{W}} f_w(x),$$

with the following four properties.

(a) $\mathcal{W}$ is a compact convex subset of some $\mathbb{R}^p$.
(b) $f_w : \mathbb{E} \to (-\infty, +\infty]$ is closed proper convex for each $w \in \mathcal{W}$, and $dom f$ is the common domain of the convex functions $f_w(x)$ with respect to x.
(c) $w \in \mathcal{W} \mapsto f_w(x)$ is closed proper concave and continuous for each $x \in dom f$.
(d) $f_w(x)$ is differentiable in x, and the derivative $(x, w) \mapsto \nabla f_w(x)$ is continuous on $dom f \times \mathcal{W}$ and differentiable in x.

Under property (c), the supremum is attained for each $x \in \mathbb{E}$. In case of a singleton set $\mathcal{W}$, this problem is one of the most studied problems in the literature, see, for instance, [21, Chapter VII] and [6, Part III]. The main difficulty in solving (Min) arises from infiniteness of the set $\mathcal{W}$.

The problem (Min) covers a quite general problem – convex conic programming, which is described in Section 1.1 below. In this paper, we aim to solve (Min) using a proximal bundle method, which is briefly interpreted in Section 1.2 below.

1.1. Convex conic programming. We consider the *convex cone program (CCP)*

$$\text{(1.1)} \qquad \begin{aligned} &\text{minimize} \quad h(x) \\ &\text{subject to} \quad \mathcal{A}(x) = b, \quad x \in \mathbb{K}, \end{aligned}$$

where h is a convex function on $\mathbb{E}$, $\mathcal{A}$ is a linear operator from $\mathbb{E}$ to $\mathbb{R}^m$, $b \in \mathbb{R}^m$, and $\mathbb{K}$ is a pointed closed convex cone of $\mathbb{E}$. The Fenchel dual is equivalent to

$$\text{(1.2)} \qquad \begin{aligned} &\text{minimize} \quad h^\sharp(z) + b^T y \\ &\text{subject to} \quad z + \mathcal{A}^*(y) \in \mathbb{K}^*, \end{aligned}$$

where $h^\sharp$, $\mathcal{A}^*$ and $\mathbb{K}^*$ denote the Fenchel conjugate of h, the adjoint of $\mathcal{A}$ and the closed dual cone $\{s \in \mathbb{E} : \langle x, s \rangle \geq 0 \; \forall x \in \mathbb{K}\}$, respectively. Note that the CCP (1.1) is equivalent to the dual (1.2) under a condition for strong duality to hold, in the sense that the values of the two objective functions coincide.

We denote by $int(\mathbb{K})$ the interior of $\mathbb{K}$. Pick an element $e \in int(\mathbb{K}^*)$, and

$$\mathcal{W}_e := \{w \in \mathbb{K} : \langle e, w \rangle = 1\},$$

which is compact. Observe that $s \in \mathbb{K}^*$ if and only if $\langle s, w \rangle \geq 0$ for all $w \in \mathcal{W}_e$; i.e., the support function $\sigma : z \mapsto \max_{w \in \mathcal{W}_e} \langle z, w \rangle$ is nonpositive at $-s$. If $e \in int(\mathbb{K}^*)$ is chosen to satisfy

(**A1**) $e = \mathcal{A}^*(\bar{y})$ for some $\bar{y} \in \mathbb{R}^m$,

then (1.2) is equivalent to the unconstrained minimization

(Min$_s$) $$\min_{z,y} \left\{ \alpha \sigma(-z - \mathcal{A}^*(y)) + h^\sharp(z) + b^T y \right\},$$

when $\alpha = \max\{0, b^T \bar{y}\}$. In Appendix A a proof of the equivalence of this minimization problem and (1.2) is presented. Assumption **A1** does hold 'for some e' when (1.1) has a bounded feasible region. Henceforth we shall assume that Assumption **A1** holds. We further assume that $dom h^\sharp$ is relatively open. This minimization problem is of the form of (Min) with $\mathcal{W}$ replaced by $\mathcal{W}_e$, and

$$f_w : (z, y) \in \mathbb{E} \times \mathbb{R}^m \mapsto \alpha \langle -z - \mathcal{A}^*(y), w \rangle + b^T y + h^\sharp(z).$$

Whether property (d) is satisfied depends on knowing the conjugate of h. There are actually many convex functions with explicit expressions of the conjugates satisfying property (d). For instance, the self-conjugate function $\frac{1}{2}|x|^2$ for $x \in \mathbb{E}$, where $|\cdot|$ denotes an arbitrary norm, the logarithmic barrier function for the positive orthant, and the logarithmic determinant function for positive definite matrices. In Section 1.1.1, we shall explain a practical application–convex quadratic conic programming, where the conjugate is known explicitly.

Let $\mathcal{O}$ and $\mathcal{O}_\star$ denote the optimal solution sets of (1.2) and (Min$_s$), respectively. If, in addition, the set $\mathcal{O}$ is nonempty, then

$$\mathcal{O}_\star = \{(z^\star, y^\star) + \tau(\mathbf{0}, \bar{y}) : (z^\star, y^\star) \in \mathcal{O}, \tau \in \mathbb{R}\}.$$

Furthermore, all feasible solutions x for (1.1) satisfy $\langle x, e \rangle = \alpha$. See Proposition A.1 in Appendix A for a proof of the above results.

1.1.1. *Convex quadratic conic programming.* Consider the *convex quadratic cone program (CQCP)*

(P) $$\begin{array}{ll} \text{minimize} & \frac{1}{2}\langle x, \mathcal{Q}(x) \rangle + \langle \hat{s}, x \rangle \\ \text{subject to} & \mathcal{A}(x) = b, \quad x \in \mathbb{K}, \end{array}$$

where $\mathcal{Q}$ is a self-adjoint positive semidefinite linear operator acting on $\mathbb{E}$, and $\hat{s} \in \mathbb{E}$. Its Fenchel dual is equivalent to the following problem, called the *dual convex quadratic cone program*,

(D) $$\begin{array}{ll} \text{minimize} & \frac{1}{2}\langle x, \mathcal{Q}(x) \rangle + b^T y \\ \text{subject to} & \mathcal{Q}(x) + \mathcal{A}^*(y) + \hat{s} \in \mathbb{K}^*, \end{array}$$

then the dual (D) is equivalent to the unconstrained minimization

$$\min_{x,y} \left\{ \alpha \sigma(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)) + b^T y + \frac{1}{2}\langle x, \mathcal{Q}(x) \rangle \right\}.$$

This minimization problem is of the form of (Min) with $\mathcal{W} = \mathcal{W}_e$, and

$$f_w : (x, y) \in \mathbb{E} \times \mathbb{R}^m \mapsto \alpha \langle -\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x), w \rangle + b^T y + \frac{1}{2}\langle x, \mathcal{Q}(x) \rangle.$$

Solving this unconstrained minimization problem solves the CQCP (P) under Assumption **A1** and a condition strong duality holds. We see that, in the case that $\alpha = 0$, this minimization problem is trivial: for any $y \in \mathbb{R}^m$, $(\mathbf{0}, y)$ is optimal if $b = \mathbf{0}$, otherwise it is unbounded. Henceforth, we only deal with the non-trivial case where $\alpha > 0$. In fact, by scaling $\mathcal{A}$ and $\mathcal{Q}$, if necessary, we shall further assume that $\alpha = 1$.

When $\mathbb{K}$ is a symmetric cone, the CQCP is called a *convex quadratic symmetric cone program (CQSCP)*. In this setting, primal-dual path-following interior-point algorithms have been provided in the literature for solving the CQSCP (P); see, for instance, [12] and [29]. The former discusses a short-step path-following algorithm for the CQSCP (P), while the latter proposes an inexact primal-dual infeasible path-following algorithm for it. Both of them present polynomial iteration bounds in terms of the rank of the underlying Euclidean Jordan algebra. The paper [12] showed that the CQSCP (P) can be reformulated as a monotone symmetric cone linear complementarity problem (SCLCP), and vice versa. Therefore, methods for solving the monotone SCLCP can be used to solve the CQSCP (P); see, for example, [32, Chapter 7] and [28]. The main event in each iteration of interior-point algorithms is to solve a search direction from a linear system of equations. Since the direct monotone SCLCP reformulation will double the dimension of the linear system, interior-point algorithms may be inefficient for large scale CQSCPs. Therefore, we resort to proximal bundle methods, which are first-order and increasingly used in many practical applications, such as in economics, optimal control and engineering; see, for example, [7, 10, 36, 42], and the references therein.

As a symmetric cone is the cone of squares of a Euclidean Jordan algebra, our approach enables one to study major classes of optimization problems (e.g., the standard convex quadratic programming and second-order cone programming) with the help of a simple and unifying Jordan-algebraic technique. When $\mathcal{Q} = \mathbf{0}$, we arrive at the class of symmetric cone programming. This has been considered in the literature, see for instance [1, 9, 37, 38]. In particular, it includes linear semidefinite programming (SDP) problems if $\mathbb{K}$ is the cone of symmetric positive semidefinite matrices. The SDPs have been exploited to develop approximation algorithms for NP-hard combinatorial optimization problems, such as the max cut problem and the travelling salesman problem. When $\mathbb{K}$ is the cone of complex Hermitian positive semidefinite matrices, it is called complex semidefinite programming (CSDP). Papers [13, 43] study CSDP and present good approximation ratios for a classical combinatorial optimization problem, known as the max-3-cut problem.

By restricting our attention to real symmetric matrices in the general formulation of CQCP in (P) and (D), we obtain a pair of primal-dual convex quadratic semidefinite programs (CQSDPs). As before, we denote the primal CQSDP by CQSDP (P). The CQSDP (P) has many practical applications in economics and engineering. It captures several well-studied problems in the literature as special cases. An example is the nearest correlation matrix problem, which arises in finance [18] and machine learning [34]. Another example is the Euclidean distance matrix completion problem (see, e.g., [4]). We refer the reader to [30] and the references therein wherein more applications of CQSDP (P) and numerical methods for solving CQSDP (P) are cited.

1.2. **The proximal bundle method.** The proximal bundle method (see, e.g., [25]) is one of the most widely studied methods for minimizing a nondifferentiable convex function φ on $\mathbb{R}^m$. It generates a sequence of *trial points* $\{x^k\}_{k=1}^{\infty}$, from which a sequence of *stability centers* $\{\hat{x}^k\}_{k=1}^{\infty}$ is derived. The sequence of trial points is recursively defined by

$$(1.3) \qquad x^{k+1} = \arg\min_{x \in \mathbb{R}^m} \{\check{\varphi}_k(x) + \frac{1}{2}\|x - \hat{x}^k\|_k^2\},$$

where

$$\check{\varphi}_k(x) = \max_{i=1,\ldots,k} \{\varphi(x^i) + \langle g(x^i), x - x^i \rangle\}$$

is a polyhedral approximation of φ, $g(x)$ is a subgradient of φ at x, and the *proximal term* $\frac{1}{2}\|x - \hat{x}^k\|_k^2$, where $\|\cdot\|_k$ is a norm on $\mathbb{R}^m$, provides stability to the algorithm. When x^{k+1} brings significant decrease in the value of φ, the algorithm will take a *serious* step (also called a *descent* step) by moving the stability center to x^{k+1}. Otherwise, a *null* step is taken, in which the stability center remains, but the underestimate $\check{\varphi}_k$ is updated to account for new information from the subdifferentials $\partial\check{\varphi}_k(x^{k+1})$ and $\partial\varphi(x^{k+1})$, which is the set $\{g \in \mathbb{R}^m : \varphi(x) \geq \varphi(x^{k+1}) + \langle g, x - x^{k+1}\rangle \text{ for all } x \in \mathbb{R}^m\}$.

In most cases, it may be tremendously expensive or even impossible to compute $\varphi(x)$ and a subgradient $g(x) \in \partial\varphi(x)$ exactly. For this reason, some modifications of traditional bundle methods involving approximate values or subgradients are proposed; for instance, [23, 26, 19, 41, 40] and [31, Section 4.5]. The proposed method in [23] is an inexact bundle method for constrained convex optimization in the presence of a Slater point, which accepts an oracle returning function and subgradient values with unknown accuracy for the function appearing in the constraints, but not for the objective function. The paper [26] presents a proximal bundle method for minimizing φ over a closed convex set in $\mathbb{R}^m$, which is on the basis of an oracle that gives a function estimate $\varphi_x \in [\varphi(x) - \epsilon_1, \varphi(x) + \epsilon_2]$ and a subgradient estimate $g_x \in \partial_{\epsilon_1+\epsilon_2}\varphi(x)$ with $\epsilon_1 + \epsilon_2 =: \epsilon$ being fixed but possibly unknown, where $\partial_\epsilon\varphi(x) = \{g \in \mathbb{R}^m : \varphi(y) \geq \varphi(x) + \langle g, y - x \rangle - \epsilon$ for all $y \in \mathbb{R}^m\}$, and both ϵ_1 and ϵ_2 are absent in the stopping criterion and descent test. The work of [26] extends that of [19] in that the latter requires exact function value. Subsequently, the authors of [33] used the introduced inexact proximal bundle method in [26] for solving the master programs of stochastic programs. Papers [40, 41] work in the setting with $\epsilon_2 = 0$ for the same problem of minimizing φ over $\mathbb{R}^m$, and use known varying tolerances ϵ_1, which is exploited in the approximation of φ and descent test of [40], and utilized in the stopping criterion and descent test of [41]. We should emphasize that a polyhedral approximation of a function has been used in [23, 26, 19, 41, 40]. In [31, Section 4.5], it also uses $\epsilon_2 = 0$ but unknown varying tolerances ϵ_1 employed in its stopping criterion and the descent test. Furthermore, it exploits a nonpolyhedral approximation for the maximum eigenvalue of a matrix-value affine mapping, which employs the semidefinite structure of the considered set. However as in [19], convergence issues are merely considered in case that the varying tolerance approaches zero.

In this paper, we propose an inexact proximal bundle method based on an oracle that, for each trial point x^k, delivers the approximate value f_{ap}^k of $f(x^k)$, an element w_{ap}^k giving rise to model updating and the error ϵ^k in the approximation employed, such that $f_{ap}^k \leq f(x^k) \leq f_{ap}^k + \epsilon^k$ and $\epsilon^k \leq \delta^{k-1}$, where δ^k is a prescribed tolerance constituting a sequence that is bounded. Therefore, it is "inexact" in the sense that it does not require to compute exact values $f(x^k)$. To produce an approximation of f in a different way than the paper [26] does, we do not use the approximate value f_{ap}^k and an approximate subgradient $\nabla f_{w_{ap}^k}(x^k)$ to construct a polyhedral approximation of f, but employ a nonpolyhedral underestimate

$$f_{\widehat{\mathcal{W}}^k} : x \mapsto \max_{w \in \widehat{\mathcal{W}}^k} f_w(x),$$

where $\widehat{\mathcal{W}}^k \subseteq \mathcal{W}$ is compact and convex, such that it is easy to find an exact solution for the following problem (1.4), and update $\widehat{\mathcal{W}}^k$ at each step to contain $\{w^{k+1}, w_{ap}^{k+1}\} \subseteq \mathcal{W}$ with the gradients $\nabla f_{w^{k+1}}(x^{k+1})$ and $\nabla f_{w_{ap}^{k+1}}(x^{k+1})$ at the next trial point x^{k+1} contained in, respectively, the subdifferential $\partial f_{\widehat{\mathcal{W}}^k}(x^{k+1})$ and approximate subdifferential $\partial_{\epsilon^{k+1}} f(x^{k+1})$, where w^{k+1} is an optimal solution of (SubProb) in Proposition 2.1 giving rise to the trial point x^{k+1}. Thus our method computes trial points as follows:

$$(1.4) \qquad x^{k+1} = \arg\min_x \{f_{\widehat{\mathcal{W}}^k}(x) + \frac{\nu_k}{2} \|x - \hat{x}^k\|_{\mathcal{M}_k}^2\},$$

where $\nu_k > 0$ is a weight controlling the effect of the proximal term, and $\|\cdot\|_{\mathcal{M}_k}$ is the norm $x \in \mathbb{E} \mapsto \sqrt{\langle x, \mathcal{M}_k x \rangle}$ with $\mathcal{M}_k : \mathbb{E} \to \mathbb{E}$ a self-adjoint positive definite linear operator. In our setting, the selection of the identity map $\mathcal{I}$ as the operator $\mathcal{M}_k$ may result in solving the subproblem (1.4) extremely expensively (see Remark 2.6), whereas the choice $\mathcal{M}_k = \mathcal{I}$ is used in [26].

1.3. **Contributions and organization.** The contribution of our results is twofold. On one hand, we extend an inexact spectral bundle method of [30] for convex quadratic semidefinite programming to the context of minimizing an unconstrained convex sup-function with an open domain, which then solves a quite general problem – the general convex conic programming.

On the other hand, when applying this method to the CQSCP, we promote the results in [30] of updating the model $\widehat{\mathcal{W}}^k$. We provide an orthonormalization process to create a system of mutually orthogonal primitive idempotents from a finite set of primitive idempotents, which includes a new approximate subgradient. We exploit this new system of primitive idempotents to generate the underlying subalgebra of $\widehat{\mathcal{W}}^{k+1}$. Our orthonormalization process can ensure that the updated model $\widehat{\mathcal{W}}^{k+1}$ possesses a desirable property to guarantee the convergence of our proposed method. Moreover, in this setting, our proposed method has the advantage of being simple and cheap in computational costs. At each iteration one only needs to compute a maximum eigenvalue inexactly and a small-sized subproblem, which is a convex quadratic cone program. It is "small-sized" in the sense that the rank of the underlying Euclidean Jordan algebra can be confined to less than 8.

The paper is organized as follows. Additional notations and terminology are presented in the rest of this section. The algorithm and its convergence analysis are provided in Section 2. This algorithm is applied to the CQSCP in Section 3, where a concrete updating rule for $\widehat{\mathcal{W}}^{k+1}$ and a routine to compute the largest eigenvalue inexactly are described. The conclusion is addressed in Section 4.

1.4. Notations and terminology.

Given a subset S of a Euclidean space, we denote by $conv(S)$ the convex hull of S.

For each positive definite m-by-m matrix $\mathbf{M}$, we denote by $\|\cdot\|_{\mathbf{M}}$ the norm $x \in \mathbb{R}^m \mapsto \sqrt{x^T \mathbf{M} x}$. The norm induced by inner product $\langle \cdot, \cdot \rangle$ of a Euclidean space $\mathbb{E}$ is denoted by $\|\cdot\|$.

Let $\mathbf{0}$ be zero element in a space or zero mapping on an algebra, whose precise meaning depends on the context. For a linear operator $\mathcal{Q}$ on $\mathbb{E}$, we denote by $\|\mathcal{Q}\|_2$ its norm $\max\{\|\mathcal{Q}(x)\| : \|x\| \leq 1\}$. The matrix representation of $\mathcal{Q}$ with respect to a given canonical orthonormal basis of $\mathbb{E}$ is denoted by $\mathbf{Q}$. Note that $\|\mathcal{Q}\|_2$ is the largest absolute value of all eigenvalues of $\mathbf{Q}$.

2. An inexact bundle method and its convergence analysis

In this section, we focus on presenting an inexact bundle method for solving (Min) in the inexact setting and analyzing its convergence. The inexactness of the algorithm lies in its capability in accepting approximate evaluations of the objective function f and approximate subgradients of f within specified tolerance.

We now introduce the main ingredients of this method.

2.1. The computation of trial points.

In order to determine a new trial point x^{k+1} from the current stability center $\hat{x}^k$, we use the model stabilized by the *Moreau-Yosida regularization*: the next trial point x^{k+1} is determined as the minimizer of the function $x \mapsto \max_{w \in \widehat{\mathcal{W}}^k} L^k(x, w)$, where L^k is the convex-concave function

$$(2.1) \qquad (x, w) \in \mathbb{E} \times \mathcal{W} \mapsto f_w(x) + \frac{\nu_k}{2}\|x - \hat{x}^k\|^2_{\mathcal{M}_k}.$$

Note that the maximum over the compact set $\widehat{\mathcal{W}}^k$ is attained since L^k is continuous.

Since the proximal term $\frac{\nu_k}{2}\|\cdot - \hat{x}^k\|^2_{\mathcal{M}_k}$ is strongly convex of modulus $\frac{\nu_k}{2}$ with respect to the norm $\|\cdot\|_{\mathcal{M}_k}$, the same holds for $L^k(\cdot, w)$ where $w \in \widehat{\mathcal{W}}^k$ is arbitrary; i.e.,

$$L^k(x, w) \geq L^k(y, w) + D_x L^k(y, w; x - y) + \frac{\nu_k}{2}\|x - y\|^2_{\mathcal{M}_k}$$

for all $x, y \in \mathbb{E}$. At any minimizer y of $x \mapsto \max_{w \in \widehat{\mathcal{W}}^k} L^k(x, w)$, the directional derivative $D_x L^k(x, \bar{w}; x - y)$, where $\bar{w} = \arg\max_{w \in \widehat{\mathcal{W}}^k} L^k(y, w)$, is zero, whence the minimizer is unique, and x^{k+1} is well-define. Moreover,

$$(2.2) \qquad L^k(x, w^{k+1}) \geq L^k(x^{k+1}, w^{k+1}) + \frac{\nu_k}{2}\|x - x^{k+1}\|^2_{\mathcal{M}_k} \quad \forall x \in \mathbb{E},$$

where $w^{k+1} \in \arg\max_{w \in \widehat{\mathcal{W}}^k} L^k(x^{k+1}, w)$.

The following proposition handles how to compute a trial point at each iteration of the algorithm.

Proposition 2.1. *The minimizer x^{k+1} of $x \mapsto \max_{w \in \widehat{\mathcal{W}}^k} L^k(x, w)$ satisfies*

$$\min_x \max_{w \in \widehat{\mathcal{W}}^k} L^k(x, w) = L^k(x^{k+1}, w^{k+1}) = \max_{w \in \widehat{\mathcal{W}}^k} \min_x L^k(x, w),$$

where w^{k+1} is an optimal solution of

(SubProb)
$$\begin{array}{ll} \text{maximize} & L^k(x_{min}(w), w) \\ \text{subject to} & w \in \widehat{\mathcal{W}}^k \end{array}$$

with $x_{min} : w \in \widehat{\mathcal{W}}^k \mapsto \arg\min_x L^k(x, w)$ being uniquely determined by

(2.3)
$$\nabla f_w(x) + \nu_k \mathcal{M}_k(x - \hat{x}^k) = \mathbf{0}.$$

Proof. Since $L^k(\cdot, w)$ is strongly convex for each $w \in \mathcal{W}$, the recession function $(L^k(\cdot, w)0^+)(y)$ is the indicator function $\delta_{\{\mathbf{0}\}}(y)$. Thus the convex functions $L^k(\cdot, w)$ for $w \in \mathcal{W}$ have no common direction of recession. Together with the boundedness of $\widehat{\mathcal{W}}^k$, it then follows from minimax theory (see, e.g., [35, Theorem 37.6]) that there exist $x^{k+1} \in \mathbb{E}$ and $w^{k+1} \in \widehat{\mathcal{W}}^k$ satisfying

$$\min_x \max_{w \in \widehat{\mathcal{W}}^k} L^k(x, w) = L^k(x^{k+1}, w^{k+1}) = \max_{w \in \widehat{\mathcal{W}}^k} \min_x L^k(x, w).$$

This w^{k+1} is then a maximizer of $w \in \mathcal{W} \mapsto \min_x L^k(x, w)$ over $\widehat{\mathcal{W}}^k$.

The strong convexity of $x \mapsto L^k(x, w)$ for each fixed $w \in \widehat{\mathcal{W}}^k$ ensures that the function $x_{min} : w \in \widehat{\mathcal{W}}^k \mapsto \arg\min_x L^k(x, w)$ is well-defined. The first order necessary and sufficient optimality condition is precisely (2.3). By property (d) of f, the continuous nonlinear system (2.3) is differentiable in x with the nonsingular Jacobian $\nabla^2 f_w + \nu_k \mathcal{M}_k$; hence it follows from the Implicit Function Theorem that x_{min} is uniquely determined by (2.3). $\qquad\square$

Remark 2.1. If we limit ourselves to the case of the CQCP (P), then the objective function resulting from (SubProb) is

$$-\frac{1}{2} \big\| \mathcal{Q}(w - \hat{x}^k) \big\|^2_{\mathcal{T}_k^{-1}} - \frac{1}{2\nu_k} \big\| b - \mathcal{A}(w) \big\|^2_{\mathbf{M}_{y,k}^{-1}} - \langle \hat{s} + \mathcal{A}^*(\hat{y}^k) + b^T \hat{y}^k + \mathcal{Q}(\hat{x}^k), w \rangle + \frac{1}{2} \langle \hat{x}^k, \mathcal{Q}(\hat{x}^k) \rangle,$$

where $\mathcal{T}_k = \mathcal{Q} + \nu_k \mathcal{M}_{x,k}$, $\mathcal{M}_{x,k} : \mathbb{E} \to \mathbb{E}$ is a given self-adjoint positive definite linear operator defining the norm $\| \cdot \|^2_{\mathcal{M}_{x,k}} = \langle \cdot, \mathcal{M}_{x,k}(\cdot) \rangle$, and $\mathbf{M}_{y,k}$ is a symmetric positive definite matrix.

Remark 2.2. In the case of convex quadratic symmetric cone programming, the subproblem (SubProb) can be rewritten as an instance of convex quadratic symmetric cone programming with a suitable choice of an approximate form $\widehat{\mathcal{W}}$ of $\mathcal{W}$. This problem can be solved efficiently by interior-point methods (for instance [29]), provided the rank of the underlying Euclidean Jordan algebra is small, say, around 8.

Lemma 2.1. *The optimal value $L^k(x^{k+1}, w^{k+1})$ of the subproblem is bounded from above by the objective value $f(\hat{x}^k)$ at the stability center $\hat{x}^k$.*

Proof. From the definition of f as a sup-function, $f(\hat{x}^k) \geq f_{w^{k+1}}(\hat{x}^k)$. From the definition of x^{k+1} as the minimizer of $L^k(\cdot, w^{k+1})$, $L^k(\hat{x}^k, w^{k+1}) \geq L^k(x^{k+1}, w^{k+1})$. The lemma then follows from the definition $L^k : (x, w) \mapsto f_w(x) + \frac{\nu_k}{2} \|x - \hat{x}^k\|^2_{\mathcal{M}_k}$. $\qquad\square$

Lemma 2.2. *The optimal solution x^{k+1} of the subproblem satisfies*

$$\sqrt{2\nu_k} \|x^{k+1} - \hat{x}^k\|_{\mathcal{M}_k} \leq \max_{w \in \mathcal{W}} \|\nabla f_w(\hat{x}^k)\|^*_k,$$

where $\|\cdot\|^*_k$ denotes the dual norm of $\sqrt{\frac{\nu_k}{2}} \|\cdot\|_{\mathcal{M}_k}$, and the maximum is attained owing to the compactness of $\mathcal{W}$ and the continuity of $w \in \mathcal{W} \mapsto \nabla f_w(\hat{x}^k)$ (see Property (d)).

Proof. The definition $L^k(x, w) = f_w(x) + \frac{\nu_k}{2}\|x - \hat{x}^k\|^2_{\mathcal{M}_k}$ and the consequence (2.2) of the strong convexity of $L^k(\cdot, w^{k+1})$ give

$$f_{w^{k+1}}(\hat{x}^k) = L^k(\hat{x}^k, w^{k+1}) \geq L^k(x^{k+1}, w^{k+1}) + \frac{\nu_k}{2}\|x^{k+1} - \hat{x}^k\|^2_{\mathcal{M}_k}$$
$$= f_{w^{k+1}}(x^{k+1}) + \nu_k\|x^{k+1} - \hat{x}^k\|^2_{\mathcal{M}_k}.$$

The convexity of $f_{w^{k+1}}$ gives $f_{w^{k+1}}(\hat{x}^k) \leq f_{w^{k+1}}(x^{k+1}) - \langle \nabla f_{w^{k+1}}(\hat{x}^k), x^{k+1} - \hat{x}^k \rangle$. Together they imply

$$\nu_k\|x^{k+1} - \hat{x}^k\|^2_{\mathcal{M}_k} \leq -\langle \nabla f_{w^{k+1}}(\hat{x}^k), x^{k+1} - \hat{x}^k \rangle \leq \sqrt{\frac{\nu_k}{2}}\|x^{k+1} - \hat{x}^k\|_{\mathcal{M}_k} \max_{w \in \mathcal{W}}\|\nabla f_w(\hat{x}^k)\|^*_k,$$

which proves the lemma. $\qquad\square$

2.2. **The optimality estimates.** The algorithm requires the existence of an oracle $\mathbf{est}_f(x, \delta)$ that generates the pair $(w, \epsilon) \in \mathcal{W} \times \mathbb{R}_+$ satisfying

(2.4)
$$f_w(x) \leq f(x) \leq f_w(x) + \epsilon \quad \text{and} \quad \epsilon \leq \delta.$$

The following theorem shows that subgradients of f_w at x are δ-subgradients of f at x.

Theorem 2.1. *For any $x \in \mathbb{E}$ and $\delta \geq 0$, there holds*

(2.5)
$$\partial_\delta f(x) \supseteq conv \bigcup_{\substack{w \in \mathcal{W} \\ f_w(x) \geq f(x) - \delta}} \partial f_w(x).$$

Proof. In light of [35, Theorem 23.5], for every $x \in \mathbb{E}$, every $w \in \mathcal{W}$ and every $x^* \in \partial f_w(x)$, we have

$$f_w(x) + f_w^\sharp(x^*) = \langle x, x^* \rangle.$$

If x further satisfies $f_w(x) \geq f(x) - \delta$ for some $w \in \mathcal{W}$ and some $\delta \geq 0$, then

$$f(x) - \delta + f^\sharp(x^*) \leq f_w(x) + f_w^\sharp(x^*) = \langle x, x^* \rangle,$$

since $f \geq f_w$ implies $f^\sharp \leq f_w^\sharp$. It then follows from Proposition 1.1 of [20] that $x^* \in \partial_\delta f(x)$. Finally, the theorem follows from the fact that $\partial_\delta f(x)$ is a convex set. $\qquad\square$

Let g^k denote the gradient $\nabla f_{w^{k+1}}(x^{k+1}) = \nu_k \mathcal{M}_k(\hat{x}^k - x^{k+1})$ involved in the first order optimality condition (2.3) for computing $x^{k+1} = x_{min}(w^{k+1})$. The convexity of $f_{w^{k+1}}$ gives

$$f_{w^{k+1}}(x) \geq f_{w^{k+1}}(x^{k+1}) + \langle g^k, x - x^{k+1} \rangle$$

for all $x \in \mathbb{E}$, whence $f(x) \geq f_{w^{k+1}}(x^{k+1}) + \langle g^k, x - x^{k+1} \rangle$ for all $x \in \mathbb{E}$. Subsequently,

(2.6)
$$f(\hat{x}^k) \leq f(x) + \langle g^k, \hat{x}^k - x \rangle + \hat{f}^k_{ap} + \hat{\epsilon}^k - f_{w^{k+1}}(x^{k+1}) - \langle g^k, \hat{x}^k - x^{k+1} \rangle,$$

where $(\hat{w}^k_{ap}, \hat{\epsilon}^k)$ is generated by $\mathbf{est}_f(\hat{x}^k, \delta^k)$ and $\hat{f}^k_{ap} := f_{\hat{w}^k_{ap}}(\hat{x}^k)$, whence

(2.7)
$$g^k \in \partial_{\tau^k + \hat{\epsilon}^k} f(\hat{x}^k), \quad \text{where } \tau^k = \hat{f}^k_{ap} - f_{w^{k+1}}(x^{k+1}) - \langle g^k, \hat{x}^k - x^{k+1} \rangle,$$

and $\tau^k + \hat{\epsilon}^k$ is nonnegative.

For simplicity of notation, we set

$$\phi^k := \nu_k\|\hat{x}^k - x^{k+1}\|_{\mathcal{M}_k}.$$

We now have the necessary ingredients for deriving the optimality estimate. In combination with the Cauchy-Schwartz inequality and the approximate subgradients in (2.7), we achieve the optimality estimates

(2.8)
$$f(\hat{x}^k) \leq f(x) + \hat{\epsilon}^k + \phi^k\|x - \hat{x}^k\|_{\mathcal{M}_k} + \tau^k$$

for all $x \in \mathbb{E}$.

The inequality (2.8) says that the point $\hat{x}^k$ is approximate optimal if the optimality measure

(2.9)
$$\varpi^k = \max\{\phi^k, \tau^k\}$$

is zero, which provides a stopping criterion for our algorithm.

2.3. An Inexact Bundle Algorithm. In this subsection, we present an inexact bundle method for the optimization problem (Min), combining ideas from [17], [26] and [31, Section 4.5].

Algorithm 2.1. *Input:* *An initial point $x^0 \in dom f$, an improvement parameter $m_L \in (0,1)$, an initial subdifferential tolerance δ^0, an initial weight $\nu_0 > 0$, a minimum weight $\nu_{min} > 0$, a weight-correction parameter $\kappa > 0$, an initial linear operator $\mathcal{M}_0$, and three stepsize factors $\alpha_1, \alpha_2, \alpha_3 > 1$. Set the step counter $k = 0$.*

1. ***(Initialization)*** *Set $\hat{x}^0 = x^0$, $i_\nu^0 = 0$. Call $[\hat{w}_{ap}^0, \hat{\epsilon}^0] = \boldsymbol{est}_f(\hat{x}^0, \delta^0)$ and set $\hat{f}_{ap}^0 = f_{\hat{w}_{ap}^0}(\hat{x}^0)$. Set $\widehat{\mathcal{W}}^0 = \{\hat{w}_{ap}^0\}$.*
2. ***(Trial point computation)*** *Solve (SubProb) to get w^{k+1}, compute x^{k+1} from (2.3).*
3. ***(Stopping criterion)*** *Compute ϖ^k from (2.9). If $\varpi^k \leq 0$, then stop.*
4. ***(Weight Correction)*** *Compute the predicted descent of f by*

$$(2.10) \qquad \vartheta^k = \hat{f}_{ap}^k - f_{w^{k+1}}(x^{k+1}),$$

   and τ^k by (2.7). If $\vartheta^k < -\kappa\tau^k$, then set $\nu_k = \frac{1}{\alpha_1}\nu_k$, $i_\nu^k = 1$, and go to Step 2.
5. ***(Descent step)*** *Call $[w_{ap}^{k+1}, \epsilon^{k+1}] = \boldsymbol{est}_f(x^{k+1}, \delta^k)$ and set $f_{ap}^{k+1} = f_{w_{ap}^{k+1}}(x^{k+1})$. If*

$$\hat{f}_{ap}^k - f_{ap}^{k+1} \geq m_L \vartheta^k,$$

   then perform the following steps. Otherwise continue with step 6.
   (i) *Set $\hat{x}^{k+1} = x^{k+1}$, $\hat{f}_{ap}^{k+1} = f_{ap}^{k+1}$, $\hat{\epsilon}^{k+1} = \epsilon^{k+1}$, $\hat{w}_{ap}^{k+1} = w_{ap}^{k+1}$, and $i_\nu^{k+1} = 0$.*
   (ii) *Select $\nu_{k+1} \in \left[\min\{\nu_{min}, \frac{1}{\alpha_2}\nu_k\}, \nu_k\right]$ and choose the operator $\mathcal{M}_{k+1}$ such that*

$$(2.11) \qquad \|x\|_{\mathcal{M}_{k+1}} \leq \|x\|_{\mathcal{M}_k} \ \forall x \in \mathbb{E}.$$

   (iii) *Pick $\delta^{k+1} > 0$ to ensure boundedness.*
   (iv) *Continue with Step 7.*
6. ***(Null step)*** *Perform the following steps.*
   (i) *Set $\hat{x}^{k+1} = \hat{x}^k$, $\hat{f}_{ap}^{k+1} = \hat{f}_{ap}^k$, $\hat{\epsilon}^{k+1} = \hat{\epsilon}^k$, $\hat{w}_{ap}^{k+1} = \hat{w}_{ap}^k$, and $i_\nu^{k+1} = i_\nu^k$.*
   (ii) *Set $\mathcal{M}_{k+1} = \mathcal{M}_k$ and $\nu_{k+1} = \nu_k$ or $\nu_{k+1} \in [\nu_k, \alpha_3\nu_k]$ if $i_\nu^{k+1} = 0$ and*

$$\hat{f}_{ap}^k - f_{ap}^{k+1} - \langle \nabla f_{w_{ap}^{k+1}}(x^{k+1}), \hat{x}^k - x^{k+1}\rangle \geq \varpi^k.$$

   (iii) *Pick $\delta^{k+1} > 0$ to ensure boundedness.*
7. ***(Updating $\widehat{\mathcal{W}}^k$)*** *Choose a compact convex set $\widehat{\mathcal{W}}^{k+1} \supseteq \{w^{k+1}, w_{ap}^{k+1}\}$.*
8. *Update $k \leftarrow k+1$ and goto step 2.*

Remark 2.3. Hereby, we derive an upper bound on the optimality measure ϖ^k, which will be used to deduce that ϖ^k can asymptotically approach zero. Using (2.7), (2.10) and the expression of ϕ^k, we see that

$$(2.12) \qquad \vartheta^k = \tau^k + \frac{1}{\nu_k}(\phi^k)^2$$

$$(2.13) \qquad = \tau^k + \frac{1}{\nu_k}\|g^k\|_{\mathcal{M}_k^{-1}}^2.$$

On the other hand, from $\tau^k + \hat{\epsilon}^k \geq 0$, we get $\tau^k \geq -\hat{\epsilon}^k$. Hence, combining with (2.12) and $\vartheta^k \geq \tau^k$, we have

$$(2.14) \qquad \varpi^k = \max\{\sqrt{\nu_k(\vartheta^k - \tau^k)}, \tau^k\},$$

$$(2.15) \qquad \varpi^k \leq \max\{\sqrt{(1+\frac{1}{\kappa})\nu_k\vartheta^k}, \vartheta^k\} \qquad\qquad \text{if } \vartheta^k \geq -\kappa\tau^k,$$

$$(2.16) \qquad \varpi^k < \sqrt{-(1+\kappa)\nu_k\tau^k} \leq \sqrt{(1+\kappa)\nu_k\hat{\epsilon}^k} \qquad\qquad \text{if } \vartheta^k < -\kappa\tau^k.$$

Using (2.13), it is clear that

$$\vartheta^k \geq -\kappa\tau^k \Leftrightarrow \frac{1}{(1+\kappa)\nu_k}\|g^k\|^2_{\mathcal{M}_k^{-1}} \geq -\tau^k \Leftrightarrow \vartheta^k \geq \frac{\kappa}{(1+\kappa)\nu_k}\|g^k\|^2_{\mathcal{M}_k^{-1}}.$$

Therefore, Step 4 will guarantee that Step 5 is entered with $\vartheta^k \geq -\kappa\tau^k$, subsequently, a descent step will bring significant decrease in the approximated value of f.

Remark 2.4. The choice of the norm $\sqrt{\frac{\nu_k}{2}}\|\cdot\|_{\mathcal{M}_k}$ is somewhat of an art, which impacts significantly both global convergence in theory and efficiency in practice. For instance, as indicated in [19], a suitable adjustment of the norm shall guide $f_{\widehat{\mathcal{W}}^k}$ to the area where it is a reliable approximation of f. To the best of our knowledge, most of the literature about the proximal bundle method selects $\mathcal{M}_k = \mathcal{I}$; see, for instance, [2, 5, 17, 19, 25, 31, 39]. In this case, there are several wise update rules published in the literature, see, for instance, [25, 39, 6]. It has been pointed out in [6, Sect.10] that a sequence of null steps between two stability centers is just for the sake of the improvement of the model but possessing no new reliable information; therefore the norm is only updated when a descent step is taken. This updating rule originates from [27], and it is also employed in [24].

Remark 2.5. Henceforth, we assume that the sequence $\{\|\mathcal{M}_k\|\}_{k=1}^{\infty}$ is bounded, any accumulation point of $\{\mathcal{M}_k\}$ is a self-adjoint positive definite linear operator. More precisely, we make the following assumption.

(**A2**) If there is an infinite number of descent steps in Algorithm 2.1, then $\{\mathcal{M}_k\}$ converges to a self-adjoint positive definite linear operator. Moreover, (2.11) is satisfied.

Remark 2.6. In practice, the choice of the operator $\mathcal{M}_k$ should make it much more efficient for solving subproblems. For instance, for the CQCP (P), one can choose the following metric which possesses the desirable property that Assumption **A2** is satisfied:

$$\mathcal{M}_k = (\mathcal{M}_{x,k}, \mathbf{M}_{y,k}) = (\mathcal{I} - \nu_k^{-1}\mathcal{Q}, I) \quad \text{and} \quad \nu_k > \|\mathcal{Q}\|_2, \quad k = 0, 1, \dots.$$

Such choice results in $\mathcal{T}_k = \nu_k\mathcal{I}$. Of course, this choice of the metric $(\mathcal{M}_{x,k}, \mathbf{M}_{y,k})$ is just an example and there is still room for improvement in the performance of the algorithm.

2.4. **Convergence analysis.** In this subsection, we will present the convergence of Algorithm 2.1.

Practical implementations may allow for $\sup \delta^k = \delta > 0$, which can only guarantee

$$(2.17) \qquad\qquad \limsup_{k\to\infty} \epsilon^k \leq \sup \epsilon^k \leq \delta,$$

and

$$(2.18) \qquad\qquad \sup \hat{\epsilon}^k \leq \sup \epsilon^{k+1} \leq \delta,$$

since the sequence $\{\hat{\epsilon}^k\}$ is derived from $\{\epsilon^{k+1}\}$. Together with (2.8) and (2.9), we see that the stability center $\hat{x}^k$ is δ- optimal(i.e., $f(\hat{x}^k) \leq \inf f + \delta$) when $\varpi^k = 0$, which justifies the stopping criterion (2.9).

We first observe that there are infinitely many loops between Steps 2 and 4 if and only if $\hat{f}_{ap}^k \leq \inf f_{\widehat{\mathcal{W}}^k} < f_{\widehat{\mathcal{W}}^k}(\hat{x}^k)$, a result which will be presented in the following theorem deducing that the current stability center $\hat{x}^k$ is δ- optimal. Meanwhile, the coming result also tackles the case of a finite number of iterations for our method. Its proof, which also depends on the above bound for $\hat{\epsilon}^k$ and Proposition 2.1, is almost identical to that in [26, Lemma 2.3], and so is omitted here.

Theorem 2.2. (1) *If* $\hat{f}_{ap}^k \leq \inf f_{\widehat{\mathcal{W}}^k}$, *then* $f(\hat{x}^k) \leq \inf f(x) + \delta$.
(2) *The optimality measure* $\varpi^k = 0$ *if and only if* $\hat{f}_{ap}^k \leq \min f_{\widehat{\mathcal{W}}^k} = f_{\widehat{\mathcal{W}}^k}(\hat{x}^k)$.
(3) *There are infinitely many loops between Steps 2 and 4 if and only if* $\hat{f}_{ap}^k \leq \inf f_{\widehat{\mathcal{W}}^k} < f_{\widehat{\mathcal{W}}^k}(\hat{x}^k)$.

From Theorem 2.2, the case of interest is that Algorithm 2.1 neither stops nor gets stuck in an infinite loop between Steps 2 and 4. Moreover, it always holds $\vartheta^k > 0$ at Step 5.

2.4.1. *Convergence in the case of finitely many descent steps.* We may first analyze the case where only finitely many descent steps occur in this sub-subsection, and will consider the case of infinite number of descent steps later. When there is only a finite number of descent steps, then there is an infinite sequence of consecutive null steps. In other words, starting with iteration K, the descent test fails in all subsequent iterations. In this case, if

$$\liminf_{k\to\infty} \nu_k = 0 \tag{2.19}$$

is satisfied, then we will see that the sequence of stability centers $\{\hat{x}^k\}$ tends to a δ- optimal solution of (Min).

Proposition 2.2. *If there is only a finite number of descent steps and* $\liminf_{k\to\infty} \nu_k = 0$, *then* $\hat{x}^k \to \hat{x}$ *and* $f(\hat{x}) \le \inf f(x) + \delta$.

Proof. Obviously, $\hat{x}^k \to \hat{x}$.

Since (2.19) is satisfied, from Steps 4 and 6 in Algorithm 2.1, one can pick a subsequence N of $\{0,1,2,\cdots\}$ such that Step 4 does decrease ν_k at iteration $k \in N$. Subsequently, via (2.16), it follows

$$\liminf_{k\to\infty} \varpi^k = 0. \tag{2.20}$$

By Algorithm 2.1, it is easy to see that $\hat{f}_{ap}^k \ge \hat{f}_{ap}^{k+1}$ for all $k = 0,1,2,\cdots$. We now show that

$$\lim_{k\to\infty} \hat{f}_{ap}^k = \inf \hat{f}_{ap}^k \le \inf f(x), \tag{2.21}$$

whence $f(\hat{x}) = \lim_{k\to\infty} f(\hat{x}^k) \overset{(2.4),(2.18)}{\le} \lim_{k\to\infty} \hat{f}_{ap}^k + \delta \overset{(2.21)}{\le} \inf f(x) + \delta$.

In view of $g^k = \nabla f_{w^{k+1}}(x^{k+1})$, it holds $f_{w^{k+1}}(x^{k+1}) + \langle g^k, x - x^{k+1}\rangle \le f_{w^{k+1}}(x) \le f(x)$ for any $x \in \mathbb{E}$. Subsequently,

$$\hat{f}_{ap}^k \le f(x) + \langle g^k, \hat{x}^k - x\rangle + \tau^k \le f(x) + \phi^k \|x - \hat{x}^k\|_{\mathcal{M}_k} + \tau^k \text{ for any } x \in \mathbb{E}. \tag{2.22}$$

Therefore, (2.21) is true by (2.9) and (2.20). $\qquad\square$

In view of Proposition 2.2, we may assume that

$$\liminf_{k\to\infty} \nu_k > 0. \tag{2.23}$$

Then the number of Step 4 that does decrease ν_k must be finite. Indeed, if Step 4 decreases ν_k for infinitely many $k \ge K$, then $\nu_k \le \frac{1}{\alpha_1}\nu_{k-1}$. However, $i_\nu^{k+1} = i_\nu^k \ne 0$ and then $\nu_{k+1} = \nu_k$ by Step 6. Subsequently, $\liminf_{k\to\infty} \nu_k = 0$ if the number of such decreases is infinite, which contradicts (2.23).

Henceforth, we shall, with no loss of generality, assume that, starting with iteration K, the descent test fails in all subsequent iterations and Step 4 does not decrease ν_k for all $k \ge K$.

We shall show that at a null step,

(1) the choice $w^k \in \widehat{\mathcal{W}}^k$ ensures that $\{L^k(x^{k+1}, w^{k+1})\}$ is monotonically increasing, and
(2) the choice $w_{ap}^k \in \widehat{\mathcal{W}}^k$ provides a lower bound on the difference $f_{w_k}(x^k) - f_{ap}^k(x^k)$ in terms of the difference $L^k(x^{k+1}, w^{k+1}) - L^{k-1}(x^k, w^k)$.

As a consequence, when an infinite sequence of consecutive null steps is encountered, the sequence $\{L^k(x^{k+1}, w^{k+1})\}$ is monotonically convergent, and we can demonstrate that the estimate $f_{w^k}(x^k)$ is asymptotically no less than the approximate value f_{ap}^k.

Lemma 2.3. *If step k is a null step (so that $\hat{x}^k = \hat{x}^{k-1}$ and $\frac{\nu_k}{2}\|\cdot\|_{\mathcal{M}_k}^2 \ge \frac{\nu_{k-1}}{2}\|\cdot\|_{\mathcal{M}_{k-1}}^2$), then*

$$\begin{aligned}
L^k(x^{k+1}, w^{k+1}) &\ge L^k(x^{k+1}, w^k) \\
&\ge L^{k-1}(x^{k+1}, w^k) \ge L^{k-1}(x^k, w^k) + \frac{\nu_{k-1}}{2}\|x^{k+1} - x^k\|_{\mathcal{M}_{k-1}}^2;
\end{aligned} \tag{2.24}$$

and

$$(2.25) \qquad f_{w^k}(x^k) \geq f_{ap}^k - 2\Delta_k - \sqrt{\Delta_k}\left(\max_{w\in\mathcal{W}}\|\nabla f_w(\hat{x}^k)\|_k^* + \max_{w\in\mathcal{W}}\|\nabla f_w(x^k)\|_{k-1}^*\right),$$

where Δ_k denotes the difference $L^k(x^{k+1}, w^{k+1}) - L^{k-1}(x^k, w^k)$, which is at least $\frac{\nu_{k-1}}{2}\|x^{k+1} - x^k\|_{\mathcal{M}_{k-1}}^2$ by (2.24).

Proof. First, we establish the chain of inequalities (2.24). The first inequality follows from the fact that w^{k+1} maximizes $L^k(x^{k+1}, \cdot)$ over $\widehat{\mathcal{W}}^k$, which contains w^k; while the second inequality follows from the assumptions $\hat{x}^k = \hat{x}^{k-1}$ and $\frac{\nu_k}{2}\|\cdot\|_{\mathcal{M}_k}^2 \geq \frac{\nu_{k-1}}{2}\|\cdot\|_{\mathcal{M}_{k-1}}^2$. The last inequality is an application of the consequence (2.2) of the strong convexity of $L^{k-1}(\cdot, w^k)$ at x^{k+1}.

Next, we establish the lower bound (2.25) by considering the two differences

$$f_{w^k}(x^k) - f_{w^{k+1}}(x^{k+1}) \quad \text{and} \quad f_{w_{ap}^k}(x^{k+1}) - f_{ap}^k.$$

Using the lower bound

$$|x - z|^2 - |y - z|^2 \geq |x - z|^2 - (|y - x| + |x - z|)^2 = -|y - x|^2 - 2|y - x||x - z|$$

for arbitrary norm $|\cdot|$ on $\mathbb{E}$, we deduce, together with the assumptions $\hat{x}^k = \hat{x}^{k-1}$ and $\frac{\nu_k}{2}\|\cdot\|_{\mathcal{M}_k}^2 \geq \frac{\nu_{k-1}}{2}\|\cdot\|_{\mathcal{M}_{k-1}}^2$,

$$\begin{aligned}
&f_{w^k}(x^k) - f_{w^{k+1}}(x^{k+1}) \\
&= L^{k-1}(x^k, w^k) - \frac{\nu_{k-1}}{2}\|x^k - \hat{x}^{k-1}\|_{\mathcal{M}_{k-1}}^2 - L^k(x^{k+1}, w^{k+1}) + \frac{\nu_k}{2}\|x^{k+1} - \hat{x}^k\|_{\mathcal{M}_k}^2 \\
&\geq L^{k-1}(x^k, w^k) - \frac{\nu_{k-1}}{2}\|x^k - \hat{x}^{k-1}\|_{\mathcal{M}_{k-1}}^2 - L^k(x^{k+1}, w^{k+1}) + \frac{\nu_{k-1}}{2}\|x^{k+1} - \hat{x}^{k-1}\|_{\mathcal{M}_{k-1}}^2 \\
&\geq -\Delta_k - \frac{\nu_{k-1}}{2}\left(\|x^{k+1} - x^k\|_{\mathcal{M}_{k-1}}^2 + 2\|x^{k+1} - x^k\|_{\mathcal{M}_{k-1}}\|x^{k+1} - \hat{x}^{k-1}\|_{\mathcal{M}_{k-1}}\right) \\
&\geq -\Delta_k - \frac{\nu_{k-1}}{2}\|x^{k+1} - x^k\|_{\mathcal{M}_{k-1}}^2 - 2\sqrt{\frac{\nu_{k-1}}{2}}\|x^{k+1} - x^k\|_{\mathcal{M}_{k-1}}\sqrt{\frac{\nu_k}{2}}\|x^{k+1} - \hat{x}^k\|_{\mathcal{M}_k}.
\end{aligned}$$

Since $f_{w_{ap}^k}$ is convex, it holds

$$\begin{aligned}
f_{w_{ap}^k}(x^{k+1}) - f_{ap}^k = f_{w_{ap}^k}(x^{k+1}) - f_{w_{ap}^k}(x^k) &\geq \langle \nabla f_{w_{ap}^k}(x^k), x^{k+1} - x^k \rangle \\
&\geq -\sqrt{\frac{\nu_{k-1}}{2}}\|x^{k+1} - x^k\|_{\mathcal{M}_{k-1}}\max_{w\in\mathcal{W}}\|\nabla f_w(x^k)\|_{k-1}^*.
\end{aligned}$$

The lower bound (2.25) is thus established by observing that
(1) the chain of inequalities (2.24) implies that the difference Δ_k is at least $\frac{\nu_{k-1}}{2}\|x^{k+1} - x^k\|_{\mathcal{M}_{k-1}}^2$,
(2) Lemma 2.2 states that the norm $\sqrt{2\nu_k}\|x^{k+1} - \hat{x}^k\|_{\mathcal{M}_k}$ is at most $\max_{w\in\mathcal{W}}\|\nabla f_w(\hat{x}^k)\|_k^*$, and
(3) w^{k+1} is the maximizer of $L^k(x^{k+1}, \cdot)$ over $\widehat{\mathcal{W}}^k$, which contains w_{ap}^k. $\qquad\square$

The above lemma, together with the boundedness of the optimal values $L^k(x^{k+1}, w^{k+1})$ (Lemma 2.1), the boundedness of the trial points (Lemma 2.2) and the continuity of $(x, w) \mapsto \nabla f_w(x)$ (Property (d)), shows that in this infinite sequence of consecutive null steps,

$$\limsup_{k\to\infty}(f_{ap}^{k+1} - f_{w^{k+1}}(x^{k+1})) \leq 0.$$

Thus, when the algorithm no longer takes descent steps, the predicted descent must eventually be zero. Indeed, every null step taken means that

$$\hat{f}_{ap}^k - f_{ap}^{k+1} < m_L(\hat{f}_{ap}^k - f_{w^{k+1}}(x^{k+1})),$$

or equivalently

$$(1 - m_L)(\hat{f}_{ap}^k - f_{w^{k+1}}(x^{k+1})) < f_{ap}^{k+1} - f_{w^{k+1}}(x^{k+1}),$$

and hence

$$0 \leq \liminf_{k \to \infty} \vartheta^k \leq \limsup_{k \to \infty} \vartheta^k \leq \limsup_{k \to \infty} \frac{1}{1 - m_L}(f_{ap}^{k+1} - f_{w^{k+1}}(x^{k+1})) \leq 0.$$

We summarize the above discussion in the following Lemma.

Lemma 2.4. *If there is only a finite number of descent steps, then $\vartheta^k \to 0$.*

Using (2.15), the above lemma shows that the optimality measure ϖ^k is eventually zero if $\limsup_{k \to \infty} \nu_k < \infty$. We will consider the case when $\limsup_{k \to \infty} \nu_k = \infty$ in the following lemma, whose proof is almost identical to that in [26, Lemma 3.5], basing on Proposition 2.1 and Lemma 2.3, and can be omitted.

Lemma 2.5. *If there is only a finite number of descent steps and $\limsup_{k \to \infty} \nu_k = \infty$, then $\liminf_{k \to \infty} \varpi^k = 0$.*

We now conclude the case when there is only a finite number of descent steps.

Proposition 2.3. *If there is only a finite number of descent steps, then $\hat{x}^k \to \hat{x}$ and $f(\hat{x}) \leq \inf f(x) + \delta$.*

Proof. Obviously, $\hat{x}^k \to \hat{x}$.

If $\liminf_{k \to \infty} \nu_k = 0$, then $f(\hat{x}) \leq \inf f(x) + \delta$ is from Proposition 2.2.

If $\liminf_{k \to \infty} \nu_k > 0$, by the proof of Proposition 2.2, it suffices to prove (2.21). Recall (2.22), then (2.21) is established by (2.9), (2.15) and Lemmas 2.4 and 2.5. $\qquad\square$

2.4.2. *Convergence in the case of infinitely many descent steps.* Now we consider the case when there is an infinite sequence $\{k_h\}$ of descent steps. At the h'th descent step, the approximated descent is at least m_L of the predicted descent $\vartheta^{k_h} := \hat{f}_{ap}^{k_h} - f_{w^{k_h+1}}(x^{k_h+1})$, so that

$$(1 - m_L)\vartheta^{k_h} \geq f_{ap}^{k_h+1} - f_{w^{k_h+1}}(x^{k_h+1})$$

Over the course of executing the algorithm, it may happen that null steps appear between two consecutive descent steps. Since we are investigating the case of an infinite number of descent steps, for the sake of convenience, we discard all null steps. In other words, we focus on the situation $\hat{x}^{k_h} = x^{k_h}$. Thus we relabel the remaining iterates and the corresponding w^{k_h} with a new index h, and assume that, for any h,

$$(2.26) \qquad x^{h+1} = x_{min}^h(w^{h+1}), \text{which solves } \nabla f_{w^{h+1}}(x) + \nu_h \mathcal{M}_h(x - x^h) = \mathbf{0},$$

$$(2.27) \qquad f_{ap}^h - f_{ap}^{h+1} \geq m_L(f_{ap}^h - f_{w^{h+1}}(x^{h+1})).$$

Moreover, at each descent step, it holds

$$(2.28) \qquad \vartheta^h = f_{ap}^h - f_{w^{h+1}}(x^{h+1}) \geq 0.$$

Recall

$$(2.29) \qquad f_{ap}^h - f_{ap}^{h+1} \geq 0;$$

i.e., the sequence $\{f_{ap}^h\}$ is monotonically decreasing.

Let $\{x^h\}_{h=1}^{\infty}$ be the sequence of stability centers specified by (2.26). We now consider two cases: 1. the sequence $\{x^h\}_{h=1}^{\infty}$ is unbounded; and 2. the sequence $\{x^h\}_{h=1}^{\infty}$ is bounded.

Proposition 2.4. *If the sequence $\{x^h\}_{h=1}^{\infty}$ is unbounded and Assumption **A2** is satisfied, then it follows $\lim_{h \to \infty} f_{ap}^h \leq \inf f(x)$, and consequently, $\limsup_{h \to \infty} f(x^h) \leq \inf f(x) + \delta$.*

Proof. Here we use a proof by contrapositive to show the first statement. Suppose that $\lim_{h \to \infty} f_{ap}^h > \inf f(x)$ is true, then

$$(2.30) \qquad \lim_{h \to \infty} \vartheta^h = 0$$

is established by (2.27), (2.28) and (2.29). Moreover, in view of $\lim_{h\to\infty} f^h_{ap} > \inf f(x)$, one can choose $\tilde{x} \in dom f$ and $\beta > 0$ such that

$$(2.31) \qquad f^h_{ap} > f(\tilde{x}) + \beta \text{ for all } h.$$

We define the weighted distance between x^{h+1} and $\tilde{x}$ as

$$d^{h+1} = \|\tilde{x} - x^{h+1}\|_{\mathcal{M}_h}.$$

Thus, we have

$$
\begin{aligned}
(d^{h+1})^2 &= \|\tilde{x} - x^h + x^h - x^{h+1}\|^2_{\mathcal{M}_h} \\
&= \|\tilde{x} - x^h\|^2_{\mathcal{M}_h} + 2\langle \tilde{x} - x^h, \mathcal{M}_h(x^h - x^{h+1})\rangle + \langle x^h - x^{h+1}, \mathcal{M}_h(x^h - x^{h+1})\rangle \\
&\overset{(2.11)}{\leq} (d^h)^2 + 2\langle \tilde{x} - x^h, \mathcal{M}_h(x^h - x^{h+1})\rangle + \langle x^h - x^{h+1}, \mathcal{M}_h(x^h - x^{h+1})\rangle.
\end{aligned}
$$

Recall that

$$g^h = \nu_h \mathcal{M}_h(x^h - x^{h+1}) \in \partial f_{\widehat{\mathcal{W}}^h}(x^{h+1}) = \nabla f_{w^{h+1}}(x^{h+1}).$$

Therefore, we obtain

$$\langle g^h, \tilde{x} - x^h\rangle \overset{(2.7)}{\leq} f(\tilde{x}) - f^h_{ap} + \tau^h \overset{(2.31)}{\leq} \tau^h - \beta.$$

Consequently, together with (2.13) and (2.27), there holds

$$(d^{h+1})^2 \leq (d^h)^2 + \frac{2}{\nu_h}(\tau^h - \beta) + \frac{2}{\nu_h}(\vartheta^h - \tau^h) = (d^h)^2 + \frac{2}{\nu_h}(\vartheta^h - \beta),$$

then, in light of (2.30), it follows $\vartheta^h - \beta \leq 0$, whence $(d^{h+1})^2 \leq (d^h)^2$ for sufficiently large h. Therefore, $\{x^h\}^\infty_{h=1}$ is bounded.

From (2.4), we have $f^h_{ap} \leq f(x^h) \leq f^h_{ap} + \epsilon^h$. In combination with (2.17) and $\lim_{h\to\infty} f^h_{ap} \leq \inf f(x)$, we deduce $\limsup_{h\to\infty} f(x^h) \leq \inf f(x) + \delta$. $\qquad\square$

At this moment, we make the following assumption:

$$(2.32) \qquad \text{there exists some } x^\circ \in dom f, \text{ such that } f^h_{ap} \geq f(x^\circ) \text{ for all } h.$$

Proposition 2.5. *If the sequence $\{x^h\}^\infty_{h=1}$ is bounded and both (2.32) and Assumption* **A2** *are satisfied, then each accumulation point of $\{x^h\}^\infty_{h=1}$ is a δ-optimal solution of* (Min).

Proof. Since the sequence of stability centers $\{x^h\}$ is bounded, it has an accumulation point, say $x^\star$. Let $\{x^{h_j}\}$ be a subsequence such that converging to $x^\star$. The continuity of f means that $\lim f(x^{h_j}) = f(x^\star)$. Using (2.29) and (2.32), we see that

$$(2.33) \qquad \lim_{h\to\infty} f^h_{ap} = \inf_h f^h_{ap} \in (-\infty, \infty).$$

It then follows from (2.27) that $\lim_{h\to\infty} \vartheta^h = 0$. Subsequently, we obtain $\lim_{h\to\infty} \varpi^h = 0$ by (2.15) and $\nu_h \leq \nu_0$. Together with (2.8), (2.17), the boundedness of $\{x^h\}$ and Assumption **A2**, it then follows that

$$f(x^\star) \leq f(x) + \delta \quad \forall x \in \mathbb{E}.$$

Therefore, any accumulation point of $\{x^h\}$ is a δ-optimal solution of (Min). $\qquad\square$

We end in subsection 2.4 with establishing global convergence of Algorithm 2.1.

Theorem 2.3. *Let $\{\hat{x}^k\}^\infty_{k=1}$ be the sequence of points generated by Algorithm 2.1. Suppose that Assumption* **A2** *holds, and the algorithm neither stops nor gets stuck in an infinite loop between Steps 2 and 4. There are two possible cases:*

(1) *If there is a finite number of descent steps, then $\hat{x}^k \to \hat{x}$, and $\hat{x}$ is a δ- optimal solution of* (Min).

(2) *If there is an infinite number of descent steps, then $\limsup_{k\to\infty} f(\hat{x}^k) \leq \inf f(x) + \delta$. In particular, each accumulation point of $\{\hat{x}^k\}^\infty_{k=1}$ is a δ-optimal solution of* (Min).

Proof. (1) It is exactly the result of Proposition 2.3.

(2) In the case of an infinite number of descent steps, without loss of generality, we can assume that for some K, $\hat{x}^k = x^k$ for any $k > K$. By the proof of Propositions 2.4 and 2.5, we only need to prove $\limsup_{k \to \infty} f(\hat{x}^k) \leq \inf f(x) + \delta$ when the assumption (2.32) is violated. Then, for any $x \in dom f$, there exists some integer $H(x)$ such that $f_{ap}^{H(x)} < f(x)$. Thus we have $\inf f_{ap}^k \leq \inf f(x)$. Consequently, for any $\varepsilon > 0$, there is some integer H such that, for all $k > H$, $f(x^k) \leq f_{ap}^k + \epsilon^k < \inf f(x) + \delta + \varepsilon$, since $\{f_{ap}^k\}_{k=1}^\infty$ is decreasing. Hence $\limsup_{k \to \infty} f(x^k) \leq \inf f(x) + \delta$. $\qquad\square$

3. Application to Convex Quadratic Symmetric Cone Programming

In this section, we consider the case when $int(\mathbb{K})$ in (P) is a symmetric cone. In this context, our adaptation is called an inexact spectral bundle method. We shall give a routine to produce a pair satisfying (2.4). Furthermore, we shall provide an explicit form for the model $\widehat{\mathcal{W}}^k$ and an orthonormalization process to update it, which is a generalization of the counterpart results in [30].

To facilitate our exposition, we list some notations in the next paragraph, which are commonly used in the setting of Euclidean Jordan algebras. The readers can find a detailed discussion on Jordan algebra theory, for instance, in [11].

Henceforth, let $(\mathbb{V}, \circ)$ be a finite-dimensional Euclidean Jordan algebra of rank $r \geq 2$ with unit e and endowed with the trace inner product $\langle \cdot, \cdot \rangle : (u, v) \in \mathbb{V} \times \mathbb{V} \to tr(u \circ v)$, where $\circ$ is the Jordan product. For any $u \in \mathbb{V}$, a natural induced norm is given by $\|u\|_F = \sqrt{tr(u \circ u)}$. We denote the eigenvalues of u in decreasing order by $\lambda_1(u) \geq \cdots \geq \lambda_r(u)$. In this sense, we define the eigenvalue map $\lambda : \mathbb{V} \to \mathbb{R}^r$ by $\lambda(x) = (\lambda_1(x), \ldots, \lambda_r(x))^T$. For a Euclidean Jordan subalgebra $\mathbb{U}$ of $\mathbb{V}$, we write $\lambda(u, \mathbb{U})$ for its eigenvalue map in $\mathbb{U}$. Likewise, we denote by $\lambda_i(u, \mathbb{U})$ the ith eigenvalue of u in the subalgebra $\mathbb{U}$, $1 \leq i \leq rk(\mathbb{U})$, where $rk(\mathbb{U})$ denotes the rank of $\mathbb{U}$. The trace of u is denoted by $tr(u)$. The inverse of u is denoted by u^{-1}. We write L_u for the *Lyapunov transformation* $L_u(v) := u \circ v$ for all $v \in \mathbb{V}$. The *quadratic representation* of u is represented by $P_u := 2L_u^2 - L_{u^2}$. Note that the trace $tr(u \circ v)$ is associative by [11, Proposition II 4.3], it follows that L_u, whence P_u, is self-adjoint under $\langle \cdot, \cdot \rangle$. We define $P_{u,v} = L_u L_v + L_v L_u - L_{u \circ v}$. Given an idempotent $c \in \mathbb{V}$, we denote by $\mathbb{V}(c, \eta)$ the eigenspace

$$\{x \in \mathbb{V} : L_c(x) = \eta x\}, \quad \eta \in \left\{1, \frac{1}{2}, 0\right\}.$$

Given a Jordan frame $\{c_1, \ldots, c_r\} \subseteq \mathbb{V}$, for $i, j \in \{1, \ldots, r\}$, we use the notations

$$\mathbb{V}_{ii} := \mathbb{V}(c_i, 1) = \{x \in \mathbb{V} : x \circ c_i = x\} = \mathbb{R}c_i,$$

$$\mathbb{V}_{ij} := \mathbb{V}(c_i, \frac{1}{2}) \cap \mathbb{V}(c_j, \frac{1}{2}) = \left\{x \in \mathbb{V} : x \circ c_i = \frac{1}{2}x = x \circ c_j\right\}, \quad i \neq j.$$

For any element $x \in \mathbb{V}$, it can be written as $x = \sum_{i=1}^r x_i c_i + \sum_{1 \leq i < j \leq r} x_{ij}$, where $x_i c_i = P_{c_i}(x) \in \mathbb{V}_{ii}$ and $x_{ij} = 4L_{c_i}(L_{c_j}(x)) \in \mathbb{V}_{ij}$.

Let $\mathcal{W}$ be the set

$$\{w \in \mathbb{K} : tr(w) = 1\}.$$

Since $\lambda_1(v) = \max_{w \in \mathcal{W}} \langle v, w \rangle$ (see, e.g., [3, Lemma 20]), the minimization problem (Min) becomes an eigenvalue minimization problem.

3.1. A routine to produce the pair.
For an element $x \in \mathbb{V}$, a triplet $(\lambda_{ap}, \tilde{c}, \epsilon)$ is said to be a *Ritz-type triplet* for $\lambda_1(x)$ if

$$(3.1) \qquad \tilde{c} \text{ is a primitive idempotent,}$$

$$(3.2) \qquad \lambda_{ap} = tr(x \circ \tilde{c}), \epsilon = \|x \circ \tilde{c} - \lambda_{ap}\tilde{c}\|_F,$$

$$(3.3) \qquad \left|\frac{\lambda_i(x) + \lambda_j(x)}{2} - \lambda_{ap}\right| \geq \lambda_1(x) - \lambda_{ap}, \ 1 \leq i < j \leq r.$$

We will see that a *Ritz-type triplet* depicts a pair of approximation eigenvalue and the corresponding idempotent for an element v of $\mathbb{V}$.

Proposition 3.1. *Let $v \in \mathbb{V}$ and $\epsilon \geq 0$. If $(\lambda_{ap}, \tilde{c}, \epsilon)$ is a Ritz-type triplet for $\lambda_1(v)$, then*

$$\lambda_{ap} \leq \lambda_1(v) \leq \lambda_{ap} + \epsilon \text{ and } \tilde{c} \in \partial_\epsilon \lambda_1(v).$$

Proof. Let $v = \sum_{i=1}^r \lambda_i(v)c_i$ be a spectral decomposition of v, and $\tilde{c} = \sum_{i=1}^r x_i c_i + \sum_{i<j} x_{ij}$ be the Pierce decomposition of $\tilde{c}$ with respect to the Jordan frame $\{c_1, \ldots, c_r\}$. From [3, Theorem 23], it holds $\lambda_{ap} = tr(v \circ \tilde{c}) \leq \lambda_1(v)$. Since $\tilde{c}$ is a primitive idempotent which has unit norm [3], $\|\tilde{c}\|_F^2 = \sum_{i=1}^r x_i^2 + \sum_{i<j} \|x_{ij}\|_F^2 = 1$.

On the other hand, we have

$$v \circ \tilde{c} - \lambda_{ap}\tilde{c} = \sum_{i=1}^r \left[\lambda_i(v) - \lambda_{ap}\right] x_i c_i + \sum_{i<j} \left[\frac{\lambda_i(v) + \lambda_j(v)}{2} - \lambda_{ap}\right] x_{ij}.$$

Therefore, we obtain

$$\epsilon = \|v \circ \tilde{c} - \lambda_{ap}\tilde{c}\|_F = \sqrt{\sum_{i=1}^r \left[\lambda_i(v) - \lambda_{ap}\right]^2 x_i^2 + \sum_{i<j} \left[\frac{\lambda_i(v) + \lambda_j(v)}{2} - \lambda_{ap}\right]^2 \|x_{ij}\|_F^2}$$

$$\geq \sqrt{\sum_{i=1}^r \left[\lambda_1(v) - \lambda_{ap}\right]^2 x_i^2 + \sum_{i<j} \left[\lambda_1(v) - \lambda_{ap}\right]^2 \|x_{ij}\|_F^2}$$

$$= |\lambda_1 - \lambda_{ap}| \sqrt{\sum_{i=1}^r x_i^2 + \sum_{i<j} \|x_{ij}\|_F^2} = \lambda_1 - \lambda_{ap},$$

where the inequality follows from (3.3). Therefore $\lambda_{ap} \leq \lambda_1(v) \leq \lambda_{ap} + \epsilon$.

In light of [3, Lemma 20], we have $\lambda_1(v) = \max_{w \in \mathcal{W}} \langle v, w \rangle = \sigma(v)$. Consequently, by Proposition 1.1 of [20] and Theorem 13.2 of [35], it is easy to see that $\tilde{c} \in \partial_\epsilon \lambda_1(v)$. $\square$

The above proposition tells us that the desired property (2.4) is satisfied by a Ritz-type triplet, which in turn delivers the desired pair.

It is known that symmetric cones are completely classified as direct sums of the following five concrete types of simple symmetric cones (see, e.g., [11, Chapter V]): (1) the second-order cone, (2) the cone of positive-definite real symmetric matrices, (3) the cone of positive-definite complex Hermitian matrices, (4) the cone of positive-definite quaternion Hermitian matrices, and (5) the exceptional 27-dimensional octonion cone. In practice, the routine $[\lambda_{ap}, \tilde{c}, \epsilon] = \mathbf{est}(x, y, \delta)$ can be handled case by case. For instance, in cases stated in Items (2) and (3), it can be implemented by the existing program **eigifp** [22], whose underlying algorithm is based on an inverse free preconditioned Krylov subspace projection method [14].

3.2. **An explicit form and updating rule for $\widehat{\mathcal{W}}^k$.** Our explicit form for $\widehat{\mathcal{W}}^k$ is inspired by the characterization of subdifferential given by the following proposition.

Proposition 3.2. *Let $v = \sum_{i=1}^r \lambda_i(v)c_i$ be a spectral decomposition of v, suppose that the multiplicity of the eigenvalue $\lambda_1(v)$ is m_v, denote by $\mathbb{V}^{(m_v)} := \mathbb{V}(c_1 + \cdots + c_{m_v}, 1)$ and $\mathbb{K}^{(m_v)} := \mathbb{V}^{(m_v)} \cap \mathbb{K}$. We have $\partial \lambda_1(v) = \{u \in \mathbb{V} : tr(u) = 1, u \in \mathbb{K}^{(m_v)}\}$.*

Proof. It is straightforward by Proposition 33 in [3], Proposition 4.1 in [16] and the statement after the proof of Corollary 34 in [3]. $\square$

In the semidefinite programming setting, the paper [17] provides the following choice for $\widehat{\mathcal{W}}^k$ at iteration k. Let $\mathbb{S}_+^n$ denote the cone of positive semidefinite matrices of order n. Let $P_k \in \mathbb{R}^{n \times r_k}$ be an orthonormal matrix, and $\overline{W}_k \in \mathbb{S}_+^n$ with $tr(\overline{W}_k) = 1$. Then

$$\widehat{\mathcal{W}}^k = \{P_k V P_k^T + \xi \overline{W}_k : tr(V) + \xi = 1, V \in \mathbb{S}_+^{r_k}, \xi \geq 0\}.$$

Motivated by this, we give our following choice for $\widehat{\mathcal{W}}^k$ in the context of symmetric cones. Suppose that $c_{1,k}, \ldots, c_{r_k,k}$ are mutually orthogonal primitive idempotents. We denote

$$e_{r_k}^k = c_{1,k} + \cdots + c_{r_k,k},$$
$$\mathbb{V}_k^{(r_k)} = \{x \in \mathbb{V} : x \circ e_{r_k}^k = x\},$$
$$\mathbb{K}^{(r_k)} = \mathbb{V}_k^{(r_k)} \cap \mathbb{K},$$

and use

(3.4) $$\widehat{\mathcal{W}}^k = \left\{ v + \xi \bar{w}_k : tr(v) + \xi = 1, v \in \mathbb{K}^{(r_k)}, \xi \geq 0 \right\}$$

as an approximation of $\mathcal{W}$ with $\widehat{\mathcal{W}}^k \subseteq \mathcal{W}$, where $\bar{w}_k \in \mathcal{W}$ is called the *aggregate subgradient*. In this subsection, we exploit aggregation to keep the rank r_k of $\mathbb{V}_k^{(r_k)}$ small. The technique is to compress the indispensable information into a single aggregate subgradient.

Let $z^{k+1} := -\hat{s} - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(x^{k+1})$. Suppose that we are given an approximate subgradient of $\lambda_1(z^{k+1})$ at iteration k, via the scheme of the proximal bundle method, we could wish that this new approximate subgradient information can be added into the updated model $\widehat{\mathcal{W}}^{k+1}$. To this end, we need the following ingredients. First, we shall present an orthonormalization process to create a system of mutually orthogonal primitive idempotents, which in turn further generate the underlying subalgebra of $\widehat{\mathcal{W}}^{k+1}$.

Let $(v^\star, \xi^\star)$ be an optimal solution for the subproblem (SubProb). Applying the spectral decomposition Theorem III.1.2 of [11] in the subalgebra $\mathbb{V}_k^{(r_k)}$ to decompose $v^\star$ into $v^\star = \sum_{i=1}^{r_k} \lambda_i^k(v^\star, \mathbb{V}_k^{(r_k)}) c_i^k$, where $\{c_1^k, \ldots, c_{r_k}^k\}$ is a Jordan frame in $\mathbb{V}_k^{(r_k)}$.

We denote $\lambda_i^k(v^\star, \mathbb{V}_k^{(r_k)})$ by λ_i^k for $i = 1, \ldots, r_k$. Since $v^\star \in \mathbb{K}^{(r_k)}$, there holds $\lambda_i^k \geq 0$, $i = 1, \ldots, r_k$. Assume that the first l_k eigenvalues are the "large" eigenvalues of $v^\star$. Let $\tilde{c}_{k+1}$ be a primitive idempotent corresponding to the approximate eigenvalue of z^{k+1}, we set

(3.5) $$\{c_{1,k+1}, \ldots, c_{r_{k+1},k+1}\} := orth\{c_1^k, \ldots, c_{l_k}^k, \tilde{c}_{k+1}\},$$
$$e_{r_{k+1}}^{k+1} = c_{1,k+1} + \cdots + c_{r_{k+1},k+1},$$
$$\mathbb{V}_{k+1}^{(r_{k+1})} = \left\{ x \in \mathbb{V} : x \circ e_{r_{k+1}}^{k+1} = x \right\},$$
$$\mathbb{K}^{(r_{k+1})} = \mathbb{V}_{k+1}^{(r_{k+1})} \cap \mathbb{K},$$

where $1 \leq r_{k+1} \leq r$, and $orth\{a_1, a_2, \ldots, a_N\}$ is the orthonormalization of a finite set of elements $\{a_1, a_2, \ldots, a_N\} \subseteq \mathbb{V}$. It is significant to note that here orthogonality is with respect to the Jordan product. In what follows, we will see that how to achieve (3.5).

The next result is crucial to orthonormalize a set of finite elements in $\mathbb{V}$.

Proposition 3.3. *For every nonzero $u \in \mathbb{V}$ and a primitive idempotent $c \in \mathbb{V}$, there exists a primitive idempotent $a \in \mathbb{V}$ such that*

$$P_u(c) = \langle u^2, c \rangle a.$$

Moreover, $P_u(c) = \mathbf{0}$ if and only if $u \circ c = \mathbf{0}$.

Proof. Since the trace is associative, $tr(P_u(c)) = \langle u^2, c \rangle$. If $P_u(c) = \mathbf{0}$, then $\langle u^2, c \rangle = 0$, therefore a can be any primitive idempotent. We are done.

Now we suppose that $P_u(c) \neq \mathbf{0}$. Using the spectral decomposition Theorems III.1.1 and III.1.2 of [11], we see that, for an element $v \in \mathbb{V}$, the following statements are equivalent: (I) v is a primitive idempotent, (II) $\lambda(v) = (1, 0, \cdots, 0)^T$, (III) v is an idempotent and $dim(\mathbb{V}(v, 1)) = 1$. By (II), together with a characterization of eigenvalues of $P_{P_u(c)}$ given by (iii) of Lemma 12 in [38], it suffices to show that $P_{P_u(c)}$ has exactly one nonzero eigenvalue.

In view of (iii) in [11, Proposition II.3.3], we have $P_{P_u(c)} = P_u P_c P_u$. Denote $dim(\mathbb{V})$ by n. By means of (iii) of [38, Lemma 12], the eigenvalues of P_c are 1 and 0. Since $\mathbb{V}(c, 1)$ is the eigenspace of P_c corresponding to the eigenvalue 1, together with (III), it deduces that P_c has

eigenvalues 1 with multiplicity one and 0 with multiplicity $n-1$. Thus, it is possible to expand the system $\{c\}$ to a canonical orthonormal (w.r.t. the trace inner product) basis $\{c, u_1, \cdots, u_{n-1}\}$ of $\mathbb{V}$, such that the matrix representation of P_c with respect to this basis is a diagonal matrix $Diag(1, 0, \cdots, 0)$. Suppose that the matrix representation of P_u with respect to this basis is A, then the matrix representation of $P_{P_u(c)}$ with respect to this basis is $ADiag(1, 0, \cdots, 0)A$. Since $u \neq \mathbf{0}$, we can assume without loss of generality that the first column of A is nonzero. Since A is symmetric, the rank of $ADiag(1, 0, \cdots, 0)A$ is one. Therefore $ADiag(1, 0, \cdots, 0)A$ has exactly one nonzero eigenvalue. The second statement follows from

$$0 = \langle u^2, c \rangle = \langle u, L_c(u) \rangle \text{ if and only if } u \circ c = \mathbf{0},$$

since L_c is self-adjoint positive semidefinite. $\qquad\square$

By adding a primitive idempotent to a finite set of mutually orthogonal primitive idempotents, a finite set of primitive idempotents is generated. However, how to orthonormalize the resulting finite set? Hereby, we propose the following algorithm to resolve this problem.

Algorithm 3.1. **Input:** *a set of linearly independent primitive idempotents $\{a_1, \ldots, a_p\}$, $2 \leq p \leq r$, where $a_1, \ldots, a_{p-1}$ are mutually orthogonal primitive idempotents.*
    Output: *a set of $l(\leq p)$ mutually orthogonal primitive idempotents*

$$\{c_1, \ldots, c_l\} := orth\{a_1, \ldots, a_p\},$$

which is a Jordan frame of the subalgebra generated by $\{a_1, \ldots, a_p\}$.
    Perform the following steps.

      1. *Set $c_i = a_i$, $i = 1, \ldots, p-1$.*
      2. *Set $c' = e - c_1 - \cdots - c_{p-1}$. If $\langle c', a_p \rangle = 0$, then $l = p-1$, stop. Otherwise, set*

$$(3.6) \quad c_p = \frac{1}{1 - \sum_{i=1}^{p-1} \langle c_i, a_p \rangle} \left[a_p + \sum_{i=1}^{p-1} \langle c_i, a_p \rangle c_i + 4 \sum_{1 \leq i < j \leq p-1} L_{c_i}(L_{c_j}(a_p)) - 2 \sum_{i=1}^{p-1} L_{c_i}(a_p) \right].$$

The following result tells us that idempotents generated by Algorithm 3.1 is well defined, which, in turn, establishes that the new approximate subgradient $\tilde{c}_{k+1}$ has the desired property

$$(3.7) \qquad\qquad\qquad\qquad \tilde{c}_{k+1} \in \mathbb{V}_{k+1}^{(r_k+1)}.$$

Theorem 3.1. *Suppose that $a_1, \ldots, a_p$ $(2 \leq p \leq r)$ are linearly independent primitive idempotents, where $a_1, \ldots, a_{p-1}$ are mutually orthogonal primitive idempotents, then*

$$\{c_1, \ldots, c_l\} := orth\{a_1, \ldots, a_p\}, \ l \leq p,$$

generated by Algorithm 3.1, is well defined, and

$$a_p \in \mathbb{V}^{(l)} = \{x \in \mathbb{V} : x \circ (c_1 + \cdots + c_l) = x\}.$$

Proof. Recall that $c' = e - c_1 - \cdots - c_{p-1}$. By virtue of Proposition 3.3, there exists a primitive idempotent $c_p \in \mathbb{V}$ such that $P_{c'}(a_p) = \langle c', a_p \rangle c_p =: \gamma c_p$. We see that $P_{c'}(a_p) = \mathbf{0}$ if and only if $\gamma = \langle c', a_p \rangle = 1 - \sum_{i=1}^{p-1} \langle c_i, a_p \rangle = 0$ by Proposition 3.3. In this case, c_p can be any primitive idempotent which is orthogonal to $c_1, \ldots, c_{p-1}$, and we set $l = p-1$.

We now assume that $\gamma \neq 0$. Since $\langle c', c_i \rangle = 0$, $P_{c'}(c_i) = \mathbf{0}$ by Proposition 3.3. Thus, $\langle c_p, c_i \rangle = \frac{1}{\gamma} \langle a_p, P_{c'}(c_i) \rangle = 0$, $i = 1, \ldots, p-1$. We set $l = p$. By [15, Proposition 6], it yields $c_p \circ c_i = \mathbf{0}$, $i = 1, \ldots, p-1$. Hence, it is possible to complete the system $\{c_1, \ldots, c_p\}$ to a Jordan frame $\{c_1, \ldots, c_p, \ldots, c_r\}$ of $\mathbb{V}$, such that

$$(3.8) \qquad\qquad\qquad\qquad c' = c_p + \cdots + c_r.$$

Using the Peirce decomposition Theorem IV.2.1 of [11], we obtain

$$(3.9) \qquad a_p = \sum_{i=1}^{p} x_i c_i + \sum_{1 \leq i < j \leq p} x_{ij} + \sum_{i=p+1}^{r} x_i c_i + \sum_{p+1 \leq i < j \leq r} x_{ij} + \sum_{1 \leq i \leq p < j \leq r} x_{ij}.$$

Applying $P_{c'}$ on both sides of (3.9), together with (3.8) and $P_{c_p+\cdots+c_r} = \sum_{i=p}^{r} P_{c_i} + 2\sum_{p \le i < j \le r} P_{c_i,c_j}$, we then have $\gamma c_p = \sum_{i=p}^{r} x_i c_i + \sum_{p \le i < j \le r} x_{ij}$. Taking the inner product with c_i on both sides of the above equation for $i = p, \ldots, r$, we have

$$(3.10) \qquad\qquad x_p = \gamma \text{ and } x_i = 0, i = p+1, \ldots, r,$$

whence $\sum_{p \le i < j \le r} x_{ij} = \mathbf{0}$, which follows $\langle a_p, c_i \rangle = 0$ by taking the inner product with c_i on both sides of (3.9), $i = p+1, \ldots, r$. Therefore, $a_p \circ c_i = \mathbf{0}$, $i = p+1, \ldots, r$ by [15, Proposition 6]. Thus, we have

$$a_p \in \left\{ \begin{array}{ll} \{x \in \mathbb{V} : x \circ (c_1 + \cdots + c_{p-1}) = x\} & \text{if } \gamma = 0, \\ \{x \in \mathbb{V} : x \circ (c_1 + \cdots + c_p) = x\} & \text{if } \gamma \ne 0. \end{array} \right.$$

Furthermore, taking the product $\circ$ with c_i on both sides of (3.9) for $i = p+1, \ldots, r$, and then summing them up on both sides, we achieve $\sum_{p+1 \le i < j \le r} x_{ij} + \frac{1}{2}\sum_{1 \le i \le p < j \le r} x_{ij} = \mathbf{0}$. On the other hand, $\sum_{p \le i < j \le r} x_{ij} = \mathbf{0}$ and $\sum_{j=p+1}^{r} x_{pj} = 2c_p \circ (\sum_{p \le i < j \le r} x_{ij}) = c_p \circ \mathbf{0} = \mathbf{0}$ imply

$$(3.11) \qquad\qquad \sum_{p+1 \le i < j \le r} x_{ij} = \mathbf{0},$$

thus

$$(3.12) \qquad\qquad \sum_{1 \le i \le p < j \le r} x_{ij} = \mathbf{0}.$$

We shall prove the equation (3.6) is valid for $\gamma \ne 0$.

In combination with (3.10), (3.11) and (3.12), we can further rewrite a_k to

$$(3.13) \qquad\qquad a_p = \sum_{i=1}^{p-1} x_i c_i + \sum_{1 \le i < j \le p-1} x_{ij} + \gamma c_p + \sum_{i=1}^{p-1} x_{ip},$$

where $x_i = \langle c_i, a_p \rangle$, $i = 1, \ldots, p-1$. Meanwhile, $1 = tr(a_p) = \sum_{i=1}^{p-1} x_i + \gamma$ implies that $\gamma = 1 - \sum_{i=1}^{p-1} \langle c_i, a_p \rangle$. Since $(\sum_{i=1}^{p-1} c_i) \circ (\sum_{1 \le i < j \le p-1} x_{ij}) = \sum_{1 \le i < j \le p-1} x_{ij}$ by [16, Proposition 4.1], we get $\sum_{i=1}^{p-1} x_{ip} = 2\left[\sum_{i=1}^{p-1} c_i \circ a_p - \sum_{i=1}^{p-1} x_i c_i - \sum_{1 \le i < j \le p-1} x_{ij} \right]$, which, combined with (3.13), gives (3.6). $\qquad\square$

We are now on a position to provide an orthonormalization procedure for (3.5).

Case 1. $c_1^k, \ldots, c_{l_k}^k, \tilde{c}_{k+1}$ are linearly dependent. Set

$$\{c_{1,k+1}, \ldots, c_{r_{k+1},k+1}\} = \{c_1^k, \ldots, c_{l_k}^k\}.$$

Case 2. $c_1^k, \ldots, c_{l_k}^k, \tilde{c}_{k+1}$ are linearly independent. In this case $\{c_{1,k+1}, \ldots, c_{r_{k+1},k+1}\}$ can be generated by Algorithm 3.1, where $r_{k+1} \le l_k + 1$.

Set $\bar{v} = \sum_{i=1}^{l_k} \lambda_i^k c_i^k$ and

$$(3.14) \qquad\qquad \bar{w}_{k+1} = \frac{\sum_{i=l_k+1}^{r_k} \lambda_i^k c_i^k + \xi^\star \bar{w}_k}{\sum_{i=l_k+1}^{r_k} \lambda_i^k + \xi^\star},$$

then $tr(\bar{w}_{k+1}) = 1$, whence $\bar{w}_{k+1} \in \mathcal{W}$.

Clearly, $\bar{v} = \sum_{i=1}^{l_k} \lambda_i^k c_{i,k+1} \in \mathbb{K}^{(r_{k+1})}$ and (3.5) implies $\tilde{c}_{k+1} \in \mathbb{V}_{k+1}^{(r_{k+1})}$ by Theorem 3.1, then $w_{ap}^{k+1} = \tilde{c}_{k+1} \in \widehat{\mathcal{W}}^{k+1}$, and $w^{k+1} = v^\star + \xi^\star \bar{w}_k = \bar{v} + (\sum_{i=l_k+1}^{r_k} \lambda_i^k + \xi^\star)\bar{w}_{k+1} \in \widehat{\mathcal{W}}^{k+1}$.

We summarize the above results as follows, which is needed to guarantee the convergence of our proposed method.

Proposition 3.4. *For* $w_{ap}^{k+1} = \tilde{c}_{k+1}$, *update formulas* (3.5) *and* (3.14) *ensure that both* w_{ap}^{k+1} *and* w^{k+1} *belong to* $\widehat{\mathcal{W}}^{k+1}$ *of the form* (3.4).

4. Conclusion

We have proposed an inexact bundle method for solving an unconstrained minimization of a convex sup-function with an open domain. The proposed algorithm finds a sequence $\{\hat{x}^k\}_{k=1}^{\infty}$. We have studied the limiting behavior of this sequence in the situation when a specified sequence of tolerance δ^k is bounded. In light of the prescribed requirements for the estimation of the value of the sup-function and the choice of the norm, we have deduced that this sequence converges to approximate solutions for the minimization problem (Min).

References

[1] F. Alizadeh and X. Yu. The $\mathcal{Q}$ method for symmetric cone programming. *J. Optim. Theory Appl.*, 149(1):102–137, 2011.

[2] P. Apkarian, D. Noll, and O. Prot. A trust region spectral bundle method for nonconvex eigenvalue optimization. *SIAM J. Optim.*, 19(1):281–306, 2008.

[3] M. Baes. Convexity and differentiablility properties of spectral functions and spectral mappings on Euclidean Jordan algebras. *Linear Algebra Appl.*, 422:664–700, 2007.

[4] M. Bakonyi and C. R. Johnson. The Euclidean distance matrix completion problem. *SIAM J. Matrix Anal. Appl.*, 16 (2):646–654, 1995.

[5] A. Belloni and C. Sagastizábal. Dynamic bundle methods. *Math. Program., Ser. A*, 120(2):289–311, 2009.

[6] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal. *Numerical Optimization. Theoretical and Practical Aspects*. Universitext. Springer, Berlin, 2nd edition, 2003.

[7] R. Borndörfer, A. Löbel, and S. Weider. A bundle method for integrated multi-depot vehicle and duty scheduling in public transit. *Lect. Notes Econ. Math. Syst.*, 600:3–24, 2008.

[8] J. M. Borwein and A. S. Lewis. *Convex analysis and nonlinear optimization. Theory and examples*. Springer, New York, second edition, 2006.

[9] C. B. Chua. The primal-dual second-order cone approximations algorithm for symmetric cone programming. *Found. Comput. Math.*, 7(3):271–302, 2007.

[10] G. Emiel and C. Sagastizábal. Incremental-like bundle methods with application to energy planning. *Comput. Optim. Appl.*, 46:305–332, 2010.

[11] J. Faraut and A. Koranyi. *Analysis on symmetric cones*. Oxford Press, New York, USA, 1994.

[12] L. Faybusovich. Euclidean Jordan algebras and interior-point algorithms. *Positivity*, 1:331–357, 1997.

[13] M. X. Goemans and D. P. Williamson. Approximation algorithms for MAX-3-CUT and other problems via complex semidefinite programming. *J. Comput. System Sci.*, 68(2):442–470, 2004.

[14] G. H. Golub and Q. Ye. An inverse free preconditioned Krylov subspace method for symmetric generalized eigenvalue problems. *SIAM J. Sci. Comput.*, 24:321–334, 2002.

[15] M. S. Gowda, R. Sznajder, and J. Tao. Some P-properties for linear transformations on Euclidean Jordan algebras. *Linear Algebra and Appl.*, 393:203–232, 2004.

[16] M. S. Gowda and J. Tao. The Cauchy interlacing theorem in simple Euclidean Jordan algebras and some consequences. *Linear Multilinear Algebra*, 59:65–86, 2011.

[17] C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. *SIAM J.Optm.*, 10(3):673–696, 2000.

[18] N. J. Higham. Computing the nearest correlation matrix—a problem from finance. *IMA J. Numer. Anal.*, 22:329–343, 2002.

[19] M. Hintermüller. A proximal bundle method based on approximate subgradients. *Comput. Optim. Appl.*, 20(3):245–266, 2001.

[20] J. B. Hiriart-Urruty. Lipschitz r-continuity of the approximate subdifferential of a convex function. *Math. Scand.*, 47(1):123–134, 1980.

[21] J. B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithm I, Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 1993.

[22] H. M. James and Q. Ye. Algorithm 845: EIGIFP: a MATLAB program for solving large symmetric generalized eigenvalue problems. *ACM Trans. Math. Softw.*, 31(2):270–279, 2005.

[23] K. C. K. C. Kiwiel and C. Lemaréchal. An inexact bundle variant suited to column generation. *Math. Program. Ser. A*, 118(1):177–206, 2009.

[24] E. Karas, A. Ribeiro, C. Sagastizábal, and M. Solodov. A bundle-filter method for nonsmooth convex constrained optimization. *Math. Program., Ser. B*, 116(1-2):297–320, 2009.

[25] K. C. Kiwiel. Proximity control in bundle mehtods for convex nondifferentiable minimization. *Math. Program.*, 46:102–122, 1990.

[26] K. C. Kiwiel. A proximal bundle method with approximate subgradient linearizations. *SIAM J. Optim.*, 16 (4):1007–1023, 2006.

[27] C. Lemaréchal and C. Sagastizábal. Variable metric bundle methods: from conceptual to implementable forms. *Math. Program.*, 76(3):393–410, 1997.

[28] G. Lesaja and C. Roos. Kernel-based interior-point methods for monotone linear complementarity problems over symmetric cones. *J. Optim. Theory Appl.*, 150(3):444–474, 2011.

[29] L. Li and K. C. Toh. A polynomial-time inexact interior-point method for convex quadratic symmetric cone programming. *J. Math-for-Ind.*, 2(B):199–212, 2010.

[30] H. L. Lin. An inexact spectral bundle method for convex quadratic semidefnite programming. *Comput. Optim. Appl.*, 53(1):45–89, 2012.

[31] S. A. Miller. *An Inexact Bundle Method for Solving Large Structured Linear Matrix Inequalities*. PhD thesis, University of California, Santa Barbara, California, 2001.

[32] Y. Nesterov and A. Nemirovskii. *Interior-Point polynomial algorithms in convex programming*. SIAM Studies in Applied Mathematics, Philadelphia, USA, 1994.

[33] W. Oliveira, C. Sagastizábal, and S. Scheimberg. Inexact bundle methods for two-stage stochastic programming. *SIAM J. Optim.*, 21(2):517–544, 2011.

[34] H.-D. Qi and D. Sun. Correlation stress testing for value-at-risk: an unconstrained convex optimization approach. *Comput. Optim. Appl.*, 45(2):427–462, 2010.

[35] R. T. Rockafellar. *Convex analysis*. Princeton University Press, Princeton, 1970.

[36] M. Rotunno and R. A. de Callafon. A bundle method for solving the fixed order control problem. In *Proc. 41st IEEE Conf. Decision and Control*, volume 3, pages 3156–3161, Las Vegas, NV, USA, 2002.

[37] S. H. Schmieta and F. Alizadeh. Associative and Jordan algebras, and polynomial time interior-point algorithms for symmetric cones. *Math. Oper. Res.*, 26(3):543–564, 2001.

[38] S. H. Schmieta and F. Alizadeh. Extension of primal-dual interior point algorithms to symmetric cones. *Math. Program.*, 96:409–438, 2003.

[39] H. Schramm and J. Zowe. A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. *SIAM J. Optim.*, 2:121–152, 1992.

[40] J. Shen, Z-Q. Xia, and L-P. Pang. A proximal bundle method with inexact data for convex nondifferentiable minimization. *Nonlinear Anal.*, 66(9):2016–2027, 2007.

[41] M. V. Solodov. On approximations with finite precision in bundle methods for nonsmooth optimization. *J. Optim. Theory Appl.*, 119(1):151–165, 2003.

[42] C. H. Teo, S. V. N. Vishwanathan, A. Smola, and Q. V. Le. Bundle methods for regularized risk minimization. *J. Mach. Learn. Res.*, 11:311–365, 2010.

[43] S. Zhang and Y. Huang. Complex quadratic optimization and semidefinite programming. *SIAM J. Optim.*, 16(3):871–890, 2006.

APPENDIX A. PROOF OF THE EQUIVALENCE BETWEEN THE CCP (1.1) AND (Min)

Proposition A.1. *Suppose that Assumption* **A1** *holds. Then the dual* (1.2) *is equivalent to a specific instance* (Min$_s$) *of* (Min). *If, in addition, the optimal solution set $\mathcal{O}$ is nonempty, then*

$$\mathcal{O}_\star = \{(z^\star, y^\star) + \tau(\mathbf{0}, \bar{y}) : (z^\star, y^\star) \in \mathcal{O}, \tau \in \mathbb{R}\}.$$

Furthermore, all feasible solutions x for the CCP (1.1) *satisfy $\langle x, e \rangle = \alpha = \max\{0, b^T \bar{y}\}$.*

Proof. Denote $g(z, y) := h^\sharp(z) + b^T y$ and $f(z, y) := \alpha\sigma(-z - \mathcal{A}^*(y)) + g(z, y)$. Let $\mathcal{F}$ and $\mathcal{F}_\star$ denote, respectively, the feasible regions of (1.2) and (Min$_s$). Then, $\mathcal{F} \subseteq \mathbb{E} \times \mathbb{R}^m = \mathcal{F}_\star$. Using Assumption **A1** and the property that $s \in \mathbb{K}^*$ if and only if $\sigma(-s) \leq 0$, we see that

$$\{(z, y) + \tau(\mathbf{0}, \bar{y}) : z \in \mathbb{E}, \tau \geq \sigma(-z - \mathcal{A}^*(y))\} \subseteq \mathcal{F},$$

whence $(z, y) + \tau(\mathbf{0}, \bar{y}) \in \mathcal{F}$ for any $(z, y) \in \mathbb{E} \times \mathbb{R}^m$ with $\tau = \sigma(-z - \mathcal{A}^*(y))$. We have proved that feasible solutions of (Min$_s$) are in one to one correspondence to those of (1.2).

Since $h^\sharp 0^+(\mathbf{0}) = 0$ by Theorem 8.5 of [35], $g0^+(\mathbf{0}, \bar{y}) = b^T \bar{y}$, whence $g0^+(\mathbf{0}, \bar{y}) = -g0^+(-\mathbf{0}, -\bar{y}) = b^T \bar{y} =: \rho$. By Theorem 8.8 of [35], it follows that

$$(A.1) \qquad g((z, y) + \tau(\mathbf{0}, \bar{y})) = g(z, y) + \tau\rho, \ \forall \ (z, y) \in \mathbb{E} \times \mathbb{R}^m, \tau \in \mathbb{R}.$$

Now we consider the cases $\rho < 0$ and $\rho \geq 0$.

<u>Case 1</u>: $\rho < 0$.

In this case, there holds $\alpha = 0$. Then $f(z, y) = g(z, y)$ for any $(z, y) \in \mathbb{E} \times \mathbb{R}^m$. By (A.1), we have

$$\lim_{\tau \to \infty} g((z, y) + \tau(\mathbf{0}, \bar{y})) = -\infty.$$

Thus, the dual (1.2) is equivalent to (Min$_s$).

<u>Case 2</u>: $\rho \geq 0$.

In this case, it yields $\alpha = \rho$. In light of (A.1) and Assumption **A1**, we get

$$(A.2) \qquad f((z, y) + \tau(\mathbf{0}, \bar{y})) = f(z, y), \ \forall \ (z, y) \in \mathbb{E} \times \mathbb{R}^m, \tau \in \mathbb{R},$$

$$(A.3) \qquad g((z, y) + \tau(\mathbf{0}, \bar{y})) = f((z, y) + \tau(\mathbf{0}, \bar{y})), \ \forall \ (z, y) \in \mathbb{E} \times \mathbb{R}^m, \tau = \sigma(-z - \mathcal{A}^*(y)).$$

Choosing $\tau_0 > 0$, we have $\sigma(-\mathcal{A}^*(\tau_0 \bar{y})) < 0$ by Assumption **A1**. It then follows from Fenchel's duality Theorem ([35, Theorem 31.1]) that

$$(A.4) \quad \inf\{h^\sharp(z) + b^T y : \sigma(-z - \mathcal{A}^*(y)) \leq 0\} = \sup_{\lambda \geq 0} \inf_{(z,y) \in \mathbb{E} \times \mathbb{R}^m} \{h^\sharp(z) + b^T y + \lambda\sigma(-z - \mathcal{A}^*(y))\}.$$

Subsequently, if $\inf\{h^\sharp(z) + b^T y : \sigma(-z - \mathcal{A}^*(y)) \leq 0\} = \inf\{h^\sharp(z) + b^T y : z + \mathcal{A}^*(y) \in \mathbb{K}^*\} = -\infty$, then $\inf_{(z,y) \in \mathbb{E} \times \mathbb{R}^m}\{h^\sharp(z) + b^T y + \alpha\sigma(-z - \mathcal{A}^*(y))\} = -\infty$. If $\inf\{h^\sharp(z) + b^T y : z + \mathcal{A}^*(y) \in \mathbb{K}^*\} > -\infty$, then (A.4) implies $\mathcal{O} \neq \emptyset$ by Theorem 4.3.8 of [8]. In this case, we shall show that

$$\mathcal{O}_\star = \{(z^\star, y^\star) + \tau(\mathbf{0}, \bar{y}) : (z^\star, y^\star) \in \mathcal{O}, \tau \in \mathbb{R}\}.$$

We first prove that if $(z^\star, y^\star) \in \mathcal{O}$, then $(z^\star, y^\star) \in \mathcal{O}_\star$. Together with (A.2), we then obtain $(z^\star, y^\star) + \tau(\mathbf{0}, \bar{y}) \in \mathcal{O}_\star$ for any $\tau \in \mathbb{R}$.

Assume for the sake of contradiction that $(z^\star, y^\star) \in \mathcal{O}$ but $(z^\star, y^\star) \notin \mathcal{O}_\star$. Then there is some $(\hat{z}, \hat{y}) \in \mathbb{E} \times \mathbb{R}^m$ such that $f(\hat{z}, \hat{y}) < f(z^\star, y^\star)$. Setting $\hat{\tau} = \sigma(-\hat{z} - \mathcal{A}^*(\hat{y}))$, we see that $(\hat{z}, \hat{y}) + \hat{\tau}(\mathbf{0}, \bar{y}) =: (\tilde{z}, \tilde{y}) \in \mathcal{F}$. For any $(z, y) \in \mathcal{F}$, we have

$$(A.5) \qquad g((z, y) + \sigma(-z - \mathcal{A}^*(y))(\mathbf{0}, \bar{y})) \leq g(z, y)$$

from (A.1). Denote $\tilde{\tau} = \sigma(-\tilde{z} - \mathcal{A}^*(\tilde{y}))$. We have

$$f((\tilde{z}, \tilde{y}) + \tilde{\tau}(\mathbf{0}, \bar{y})) \overset{(A.3)}{=} g((\tilde{z}, \tilde{y}) + \tilde{\tau}(\mathbf{0}, \bar{y})) \overset{(A.5)}{\leq} g(\tilde{z}, \tilde{y}) \overset{(A.3)}{=} f((\hat{z}, \hat{y}) + \hat{\tau}(\mathbf{0}, \bar{y}))$$

$$\overset{(A.2)}{=} f(\hat{z}, \hat{y}) < f(z^\star, y^\star) \leq g(z^\star, y^\star),$$

which contradicts $(z^\star, y^\star) \in \mathcal{O}$.

Similarly, we can show that if $(z^\star, y^\star) \in \mathcal{O}_\star$, then $(z^\star, y^\star) + \tau^\star(\mathbf{0}, \bar{y}) \in \mathcal{O}$ with $\tau^\star = \sigma(-z^\star - \mathcal{A}^*(y^\star))$. Therefore, for any $(z^\star, y^\star) \in \mathcal{O}$, there holds $(z^\star, y^\star) \in \mathcal{O}_\star$, whence $(z^\star, y^\star) + \tau^\star(\mathbf{0}, \bar{y}) \in \mathcal{O}$. Then

$$g(z^\star, y^\star) = g((z^\star, y^\star) + \tau^\star(\mathbf{0}, \bar{y})) \overset{(A.3)}{=} f((z^\star, y^\star) + \tau^\star(\mathbf{0}, \bar{y})) \overset{(A.2)}{=} f(z^\star, y^\star).$$

We conclude that the dual (1.2) is equivalent to (Min$_s$).

If x is feasible for the CCP (1.1), then $\mathcal{A}(x) = b$. Hence

$$0 = \langle \mathcal{A}(x) - b, \bar{y} \rangle = \langle x, \mathcal{A}^*(\bar{y}) \rangle - b^T \bar{y} = \langle x, e \rangle - b^T \bar{y}.$$

Noting that $\langle x, e \rangle \geq 0$, we achieve $\langle x, e \rangle = b^T \bar{y} = \alpha$. $\qquad\square$

($*$) Division of Mathematical Sciences, School of Physical & Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore

E-mail address: cbchua@ntu.edu.sg

($\dagger$) School of Mathematics and Computer Science, Fujian Normal University, Fuzhou 350007, PR China

E-mail address: LINH0016@e.ntu.edu.sg, linhuiling380@hotmail.com