The Inexact Spectral Bundle Method for Convex Quadratic Semidefinite Programming

HUILING LIN[†,‡]

Abstract. We present an inexact spectral bundle method for solving convex quadratic semidefinite optimization problems. This method is a first-order method, hence requires much less computational cost each iteration than second-order approaches such as interior-point methods. In each iteration of our method, we solve an eigenvalue minimization problem inexactly, and solve a small convex quadratic semidefinite programming as a subproblem. We give a proof of the global convergence of this method using techniques from the analysis of the standard bundle method, and provide a global error bound under a Slater type condition for the problem in question. Numerical experiments with matrices of order up to 3000 are performed and the computational results establish the effectiveness of this method.

1. Introduction

1.1. **Overview.** Let $\mathcal{S}^n$ denote the space of $n \times n$ symmetric matrices endowed with the standard trace inner product $\langle X, Y\rangle = tr(XY)$ for $X, Y \in \mathcal{S}^n$, and $X \succeq 0$ means that X is positive semidefinite. The convex quadratic semidefinite programming (P) satisfies the following

$$\text{(P)} \qquad \begin{array}{ll} \min & \langle X, \mathcal{Q}(X)\rangle/2 + \langle C, X\rangle =: h(X) \\ \text{s.t.} & \mathcal{A}(X) = b \\ & X \succeq 0, \end{array}$$

where $\mathcal{Q}$ is a self-adjoint positive semidefinite linear operator acting on $\mathcal{S}^n$, $\mathcal{A}$ is a linear operator from $\mathcal{S}^n$ to $\mathbb{R}^m$ and $b \in \mathbb{R}^m$, $C \in \mathcal{S}^n$. Let $\mathcal{A}^*$ be the adjoint of $\mathcal{A}$. The Lagrange dual is equivalent to the following problem, called dual convex quadratic semidefinite programming (D),

$$\text{(D)} \qquad \begin{array}{ll} \min & \langle X, \mathcal{Q}(X)\rangle/2 + b^T y =: g(X; y) \\ \text{s.t.} & Z = \mathcal{Q}(X) + \mathcal{A}^*(y) + C \\ & Z \succeq 0. \end{array}$$

It is noted that (P) is a special case of the so-called convex quadratically constrained quadratic semidefinite programs (CQCQSDPs) proposed by Sun and Zhang in [46], where the constraints are quadratic and convex.

Under some mild assumptions, in Section 2, we show that (D) is equivalent to, denoted by eigenvalue minimization problem (EigForm),

$$\text{(Eigform)} \qquad \min_{X,y} \ a\lambda_1(-C - \mathcal{A}^*(y) - \mathcal{Q}(X)) + b^T y + \langle X, \mathcal{Q}(X)\rangle/2 =: f(X; y),$$

where $\lambda_1(-Z)$ is the largest eigenvalue of $-Z = -C - \mathcal{A}^*(y) - \mathcal{Q}(X)$. Our proposed method is aimed at solving this eigenvalue minimization problem.

The problem (P) has many practical applications in economics and engineering. It captures several well-studied problems in the literature as special cases. An underlying

2000 *Mathematics Subject Classification.* 90C25, 34A45, 65G99.

Key words and phrases. Semidefinite programming, nonsmooth optimization methods, inexact spectral bundle method, eigenvalue minimization problem, approximate subgradients.

example of (P) is the nearest correlation matrix problem, where given a data matrix $B \in \mathcal{S}^n$ and a self-adjoint operator $\mathcal{L}$ on $\mathcal{S}^n$, one wants to solve

$$
\begin{aligned}
&\min \ \|\mathcal{L}(X-B)\|^2/2 \\
&s.t.\ X_{ii}=1, i=1,\ldots,n \\
&\quad X \succeq 0.
\end{aligned}
\tag{1.1}
$$

It arises from the finance [15] and machine learning [41]. Various methods have been developed for solving the nearest correlation matrix problem. Very recent works include [10, 40] and the references therein. The former presented a second-order algorithm—an inexact smoothing Newton method—to solve it and demonstrated high efficiency of the proposed method in numerical experiments. Incidentally, the second-order techniques could be introduced to bundle methods, see, e.g., [38]. Such methods have high accuracy but high computational cost. The latter used an augmented Lagrangian dual-based approach which is quadratically convergent. Another example is the Euclidean distance matrix completion problem (see, e.g., [3]). For more applications of the convex quadratic semidefinite programming, we refer the readers to [49] and to the references therein.

The problem (P) can be solved by several existing methods. It can be reformulated as either a standard semidefinite-quadratic-linear program (see [51]) by introducing a few additional linear constraints and variables, or a semidefinite linear complementarity problem (see [26]) via exploiting its KKT system. However, their computational cost leaves a lot to be desired for large scale problems. Therefore it is indispensable to design algorithms specifically for large scale (P) which are amenable to take advantage of the specific structure of the problem. To the best of our knowledge, there are so far just a few such methods. A theoretical primal–dual potential reduction algorithm using NT directions was obtained for solving (P) in [37]. The authors suggested to use the conjugate gradient method to compute an approximate search direction. In [49], the author proposed an inexact primal–dual path-following method with three classes of preconditioners for the augmented equation via the preconditioned symmetric quasi-minimal residual (PSQMR) iterative solver. It converges quickly under suitable nondegeneracy assumptions. It is this algorithm that we are going to compare with. In addition, this algorithm was adopted to obtain a positive semidefinite correlation matrix in [9] and its polynomial iteration complexity was established in [29]. In her thesis [53], Zhao designed a semismooth Newton-CG augmented Lagrangian method for large scale (P). It is noted that all these methods are not first-order methods.

More recently, a modified alternating direction method for the more general case CQCQSDPs was proposed in [46]. It is a first-order method. Its main idea was to reformulate CQCQSDPs as a variational inequality problem and then apply the alternating direction method to the reformulated problem. The authors showed the global convergence and provided numerical evidence to show the effectiveness of this method.

As mentioned in [46], first-order methods usually require much less computation per iteration, and therefore might be suitable for relatively large problems. Meanwhile, this type of methods is relatively easy to implement but at the cost of a poor convergence rate, which will be illustrated by the performance results of our proposed algorithm on the nearest correlation matrix problem.

In this paper, we propose another very different first-order method—the inexact spectral bundle method— to solve (P) . Our proposed method has the advantage of being simple and cheap in computation. At each iteration one only needs to compute a maximum eigenvalue and a small-sized (less than 30) subproblem, which is a convex quadratic semidefinite programming. A major difference between our method and the spectral bundle method for the linear semidefinite programming [14] is that the maximum eigenvalue problem is no longer sparse in the case where both the linear map $\mathcal{A}$ and the matrix C are sparse. To avoid the high cost of computing the maximum eigenvalue, we use an existing program **eigifp** [20] to compute it inexactly. Meanwhile, the subproblems can be solved efficiently by an interior-point algorithm.

The spectral bundle method is a specialization of the proximal bundle method of Kiwiel [24] to the largest eigenvalue optimization problem, which is one of three popular types of first-order bundle methods. We refer the readers to [47] for realizing the classification of bundle methods.

We do not review the state-of-the-art of bundle methods but are inclined to provide some elementary references here. The work [32] is an excellent primer and [5, 18] are complete treatment of the subject. Furthermore, the work [31] addressed a complete development and history of bundle methods. Owing to the fact that interior-point algorithms perform poorly with large scale problems because of their high demand for storage and being time-consuming, there has been a recent, renewed interest in bundle methods. There are very recent related works on this subject presenting in, for instance, [2, 4, 21, 23, 25, 36, 45] and the references therein. We emphasize that a bundle method has been employed to solve a quite general problem—the equilibrium problem in [36], which covers a wide range, such as the optimization problem, the variational inequality problem, the Nash equilibrium problem in noncooperative games, the fixed point problem, the nonlinear complementarity problem and the vector optimization problem. At the same time, they are increasing used in many practical applications, for instance, in economics, optimal control and engineering, see, for example, [6, 7, 42, 47] and the references therein.

In this paper, we would like to make the following assumptions:

(A1) The problem (P) is strictly feasible.

(A2) There exists some $\bar{y} \in \mathbb{R}^m$ such that $I = \mathcal{A}^*(\bar{y})$.

The assumption **(A1)** provides the sufficient condition for the strong duality to hold, while **(A2)** is a bridge that will link (D) with (EigForm).

The essential structure of our proposed algorithm is that of Algorithm 4.1 in [14]. Our proposed algorithm generates a sequence of *trail points* $\{(X^k; y^k)\}_{k=1}^{\infty}$ that contains a subsequence of the so called *stability centers* $\{(\widehat{X}^k; \hat{y}^k)\}_{k=1}^{\infty}$, starting from $(X^0; y^0)$ and defined by

$$(1.2) \qquad (X^{k+1}; y^{k+1}) = \min_{X,y} \left\{ f_{\widehat{\mathcal{W}}^k}(X; y) + u_k(\|X - \widehat{X}^k\|^2_{\mathcal{M}_{X,k}} + \|y - \hat{y}^k\|^2_{\mathbf{M}_{y,k}})/2 \right\},$$

where $f_{\widehat{\mathcal{W}}^k}(X; y)$ is an approximation of $f(X; y)$, see Section 3.1 for its definition, the weight $u_k > 0$ controls the effect of the proximity of the next iterate, $\mathcal{M}_{X,k} : \mathcal{S}^n \to \mathcal{S}^n$ is a given self-adjoint positive definite linear operator such that $\|\cdot\|^2_{\mathcal{M}_{X,k}} = \langle \cdot, \mathcal{M}_{X,k}(\cdot)\rangle$, and $\mathbf{M}_{y,k}$ is a positive definite matrix. When $(X^{k+1}; y^{k+1})$ is a good trial point in the sense that it brings significant descent for f, we shall say that a *serious* step or *descent* step is made. Otherwise, a *null* step occurs. As has been pointed out in [33], the underlying structure of bundle methods is that of the approximate proximal point method in Algorithm 4.4.1 in [33]. What is more, [43] demonstrated that bundle methods for computing zeroes of general maximal monotone operators can be cast as a special case in a certain class of hybrid proximal point algorithms, and proved the linear rate of convergence for bundle methods. To be brief, bundle methods in the literature combine the cutting plane method and Moreau-Yosida regularization technique, which guarantee them to possess both descent and stability properties. They are nonsmooth optimization methods.

In this paper, our proposed method is a twofold generalization of the spectral bundle method of [14]. First, it extends to the inexact setting. Our method can be viewed as a dual approach in the sense that we will solve the equivalent form of the dual problem under reasonable assumptions. We will see that in Section 2 (D) can be reformulated as an eigenvalue optimization problem (Eigform). We show that a sequence of *stability centers* $\{(\widehat{X}^k; \hat{y}^k)\}_{k=1}^{\infty}$ generated by our proposed algorithm either is unbound or converges to an exact solution for the reformulated problem, if the eigenvalue tolerance is allowed to approach to zero; while it converges to an approximate solution when the eigenvalue tolerance is trend to a positive scalar. However, this is not the end of the story. Our proposed method obtains a sequence of points as a byproduct, whose accumulation point is an exact or approximate solution for (P) . Meanwhile, we obtain a global error bound for (Eigform) under some mild conditions.

Second, we extend the domain of f to the $\mathcal{S}^n \times \mathbb{R}^m$ setting, while [14] just considered the case when f defined on $\mathbb{R}^m$. It is the matrix variable that makes a key difference to choose $(\mathcal{M}_{X,k}, \mathbf{M}_{y,k})$. Theoretically, we establish that our proposed algorithm converges globally with the hypothesis that

$$\|X\|_{\mathcal{M}_{X,k+1}} \leq \|X\|_{\mathcal{M}_{X,k}} \text{ and } \|y\|_{\mathbf{M}_{y,k+1}} \leq \|y\|_{\mathbf{M}_{y,k}} \text{ for any } X \in \mathcal{S}^n, y \in \mathbb{R}^m \tag{1.3}$$

at a descent step, while

$$\|X\|_{\mathcal{M}_{X,k+1}} \geq \|X\|_{\mathcal{M}_{X,k}} \text{ and } \|y\|_{\mathbf{M}_{y,k+1}} \geq \|y\|_{\mathbf{M}_{y,k}} \text{ for any } X \in \mathcal{S}^n, y \in \mathbb{R}^m \tag{1.4}$$

at a null step. However, from the computational cost view, we expect that the choice should make the subproblems more tractable. For instance, $(\mathcal{M}_{X,k}, \mathbf{M}_{y,k}) = (\mathcal{I}, I)$ is accepted in theory, since it possesses the properties (1.3) and (1.4). However, in general, one may not solve the subproblems efficiently with this choice, regardless of how robust the strategy is for modifying u_k. The reason is that such choice results in computing the inverse of the operator $\mathcal{Q} + u_k\mathcal{I}$, whose computational cost may be the same as that of solving the minimum eigenvalue problem (Eigform) for large-scale problems. In Section 4, we provide an efficient choice of $(\mathcal{M}_{X,k}, \mathbf{M}_{y,k})$.

The paper is organized as follows. Some basic notation and terminology are presented in the rest of this section. The relation between the convex quadratic semidefinite programming and the eigenvalue minimization problem is explored in Section 2. The algorithm and its convergence analysis are provided in Section 3. Meanwhile, a global error bound under mild assumptions is given in the same section. Numerical examples for the nearest correlation matrix problem and some numerical comparison results are stated in Section 4. The conclusion and some possible improvements are addressed in Section 5.

1.2. **Basic notation and terminology.** At this point, we need to recall some notation. Throughout this paper, $\bar{n} = n(n+1)/2$. Given a subset T of an Euclidean space and a scalar c, we denote by cT the set $\{ct | t \in T\}$. The notation $B(x, \gamma)$ denotes the open ball in an Euclidean space centered at x with radius γ.

For given $x, y \in \mathbb{R}^m$, the inner product is $\langle x, y \rangle = x^T y$. The Euclidean norm is defined by $\|x\| = \sqrt{x^T x}$. Let $\|x\|_M = \sqrt{x^T M x}$ be a norm on $\mathbb{R}^m$ with M being a positive definite matrix. We let e be a vector of all ones of appropriate dimension, and e_i be a vector of appropriate dimension with a 1 in the i-th coordinate and $0'$s elsewhere. For a finite-valued convex function f, we denote the subdifferential and the ϵ-subdifferential of f at $x \in \mathbb{R}^m$ by $\partial f(x) = \{g \in \mathbb{R}^m | f(y) \geq f(x) + \langle g, y - x \rangle \text{ for all } y \in \mathbb{R}^m\}$ and $\partial_\epsilon f(x) =$

$\{g \in \mathbb{R}^m | f(y) \geq f(x) + \langle g, y-x \rangle - \epsilon \text{ for all } y \in \mathbb{R}^m\}$ respectively. We define the distance from a point x to the set S in an Euclidean space as

$$d(x,S) = \begin{cases} \inf_{y \in S} \|x-y\| & \text{if } S \text{ is nonempty,} \\ +\infty & \text{otherwise} \end{cases}$$

where $\|\cdot\|$ denotes an appropriate norm.

Let $\mathbb{R}^{m\times n}$ be the space of $m \times n$ matrices and $\mathcal{S}^n$ be the space of symmetric matrices of order n. The notations $\mathcal{S}^n_+$ and $\mathcal{S}^n_-$ refer to cone of positive semidefinite and cone of negative semidefinite matrices, respectively. The matrix $X - Y \in \mathcal{S}^n$ being positive semidefinite (resp. positive definite) denotes by $X \succeq Y$(resp. $X \succ Y$). We denote I by an identity matrix of appropriate order. For a given $X = (X_{ij})_{n\times n} \in \mathcal{S}^n$, its trace is denoted by $tr(X) = \sum_{i=1}^n X_{ii}$, and its Frobenius norm is $\|X\| = \sqrt{\langle X, X\rangle}$. Its eigenvalues are arranged in decreasing order, denoted by $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$. In this sense, we define the eigenvalue map $\lambda : \mathcal{S}^n \to \mathbb{R}^n$ by $\lambda(X) = (\lambda_1(X), \cdots, \lambda_n(X))^T$. For given $X, Y \in \mathcal{S}^n$, we denote the inner product by $\langle X, Y\rangle = tr(XY)$. The notation $Diag(\mu)$ is the diagonal matrix with diagonal entries $\mu_1, \cdots, \mu_n$.

Let $\mathcal{A}^* : \mathbb{R}^m \to \mathcal{S}^n$ be the adjoint operator of the linear operator $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ defined by, for all $(X; y) \in \mathcal{S}^n \times \mathbb{R}^m$, $\langle X, \mathcal{A}^*(y)\rangle = y^T \mathcal{A}(X)$. When $\mathcal{A}^*(y) = \sum_{i=1}^m y_i A_i$ with given $A_1, A_2, \cdots, A_m \in \mathcal{S}^n$, it holds $\mathcal{A}(X) = (\langle A_1, X\rangle, \langle A_2, X\rangle, \cdots, \langle A_m, X\rangle)^T$. Let $\mathcal{I} : \mathcal{S}^n \to \mathcal{S}^n$ be the identity map. A linear operator $\mathcal{Q} : \mathcal{S}^n \to \mathcal{S}^n$ is called self-adjoint positive semidefinite if for all $X, Y \in \mathcal{S}^n$, $\langle \mathcal{Q}(X), Y\rangle = \langle X, \mathcal{Q}(Y)\rangle$ and $\langle \mathcal{Q}(X), X\rangle \geq 0$. We define $\|\mathcal{Q}\| = \max\left\{\|\mathcal{Q}(X)\| \middle| \|X\| \leq 1\right\}$. The matrix representation of $\mathcal{Q}$ with respect to a given canonical orthonormal basis of $\mathcal{S}^n$ is denoted by $\mathrm{mat}(\mathcal{Q})$. It is noted that $\|\mathcal{Q}\|$ is the largest absolute values of all eigenvalues of $\mathrm{mat}(\mathcal{Q})$. We denote $\|X\|_{\mathcal{M}} = \sqrt{\langle X, \mathcal{M}(X)\rangle}$ by a norm on $\mathcal{S}^n$ with $\mathcal{M}$ being a self-adjoint positive definite linear operator.

2. The Lagrange dual and eigenvalue optimization

In this section, we shall formulate the dual convex quadratic semidefinite programming to an eigenvalue minimization problem with the help of **(A2)**.

In (D), $Z \succeq 0$ is equivalent to $0 \geq -\lambda_n(Z) = \lambda_1(-Z)$. Thus after adding it into the objective function by a Lagrange multiplier $a \geq 0$, we have the form of (EigForm). The assumption **(A2)** allows a reformulation of (D) as (EigForm), which will be explained in the following proposition. The proof of this proposition shares the same argument in the proof for Proposition 5.1.1 in [13]. Its proof is omitted here.

Proposition 2.1. *If* **(A2)** *satisfies , let* $a = \max\{0, b^T\bar{y}\}$*, then (D) is equivalent to (Eigform). Furthermore, if* X *is a feasible solution for (P), then* $tr(X) = a$*.*

Here we give an example to illustrate the first part of the above result. we shall see that the nearest correlation matrix problem has properties **(A1)** and **(A2)**.

Example 2.1. *Now we consider the nearest correlation matrix problem. The problem (P) resulting from Problem (1.1) has* $\mathcal{Q} = \mathcal{L}^2$, $C = -\mathcal{L}^2(B)$, $b = e$ *and* $\mathcal{A}(X) = (\langle A_1, X\rangle, \cdots, \langle A_n, X\rangle)^T$*, where* $A_i = e_i e_i^T$, $i = 1, \cdots, n$*. Thus* $\mathcal{A}^*(y) = Diag(y)$ *for any* $y \in \mathbb{R}^n$*. Moreover Problem (1.1) satisfies* **(A1)** *and* **(A2)***. Indeed,* I *is strictly feasible for Problem (1.1) and* $\mathcal{A}^*(e) = I$*. By Proposition 2.1, the equivalent form of (Eigform) resulting from Problem (1.1) is*

$$\min_{X,y} \quad n\lambda_1(\mathcal{L}^2(B) - Diag(y) - \mathcal{L}^2(X)) + e^T y + \langle X, \mathcal{L}^2(X)\rangle/2, \tag{2.1}$$

which says that the assumptions **(A1)** *and* **(A2)** *are reasonable in reality.* □

By virtue of Proposition 2.1, we see that, in the case that $a = 0$, (Eigform) is trivial: for any $y \in \mathbb{R}^m$, $(0; y)$ is optimal if $b = 0$, otherwise it is unbounded. Therefore, the case that $a > 0$ is a question of our interest. Henceforth, we suppose that $a > 0$.

We now are in a position to consider the subdifferential of f.

Applying a classical chain rule [18, Theorem XI.3.2.1], we obtain the expression of the subdifferential of f at $(X; y) \in \mathcal{S}^n \times \mathbb{R}^m$

$$\partial f(X; y) = \left\{(\mathcal{Q}(X - W); b - \mathcal{A}(W))\middle| W \in a\partial\lambda_1(-C - \mathcal{A}^*(y) - \mathcal{Q}(X))\right\}. \tag{2.2}$$

Note that we here follow the MATLAB convention and use ";" for adjoining scalars, vectors, or matrices in a column.

Alternatively, $f(X, y)$ may be expressed in terms of composite functions. More precisely, let $\Phi : \mathcal{S}^n \times \mathbb{R}^m \to \mathcal{S}^n$ be defined by

$$\Phi(X; y) = a(-C - \mathcal{A}^*(y) - \mathcal{Q}(X)) + (\langle X, \mathcal{Q}(X)\rangle/2 + b^T y)I,$$

and $p : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz mapping, by choosing $p(x_1, \cdots, x_n) = x_1$, we obtain $f(X; y) = (p \circ \lambda \circ \Phi)(X; y)$, where the binary operation $\circ$ is the composition of functions. Such reformulation of $f(X; y)$ leads to an alternate description of $\partial f(X; y)$. We will see that such description contributes to the error bound analysis in Section 3.

Plainly, Φ is everywhere Fréchet-differentiable on $\mathcal{S}^n \times \mathbb{R}^m$. For given $(X; y) \in \mathcal{S}^n \times \mathbb{R}^m$, let

$$J\Phi(X; y) : \mathcal{S}^n \times \mathbb{R}^m \to \mathcal{S}^n,$$

a linear mapping, denote the derivative of Φ at $(X; y)$, and

$$J\Phi(X; y)^* : \mathcal{S}^n \to \mathcal{S}^n \times \mathbb{R}^m$$

be the adjoint of $J\Phi(X; y)$.

Via the definition of Fréchet-differentiability, we have, for any $(D; d) \in \mathcal{S}^n \times \mathbb{R}^m$,

$$J\Phi(X; y)(D; d) = (b^T d + \langle \mathcal{Q}(D), X\rangle)I - a\mathcal{A}^*(d) - a\mathcal{Q}(D),$$

hence, for any $G \in \mathcal{S}^n$,

$$J\Phi(X; y)^*(G) = (tr(G)\mathcal{Q}(X) - a\mathcal{Q}(G); tr(G)b - a\mathcal{A}(G)).$$

Here is the alternate description of $\partial f(X; y)$.

Theorem 2.1. *For a given $(X; y) \in \mathcal{S}^n \times \mathbb{R}^m$, it holds*

$$\begin{aligned}\partial f(X; y) \subseteq \Big\{J\Phi(X; y)^*(G)\Big| \exists\, G \in \partial(p \circ \lambda)(\Phi(X; y)) \text{ and } \mu \in \partial p(\lambda(\Phi(X; y))) \text{ s.t.} \\ G = U^T Diag(\mu)U, \text{ with } U \text{ orthogonal and} \\ U^T(Diag(\lambda(\Phi(X; y))))U = \Phi(X; y)\Big\}.\end{aligned}$$

Proof. The proof is straightforward by chain rule (see e.g., [35, Corollary 6.3]) and [28, Theroem 6]. □

The rest of this section is to review some characterizations of the maximum eigenvalue of a symmetric matrix.

For any matrix $A \in \mathcal{S}^n$, using *Rayleigh's variational formulation*, we have

$$\lambda_1(A) = \max_{v \in \mathbb{R}^n, \|v\|=1} \left\{v^T A v\right\} = \max_{v \in \mathbb{R}^n, \|v\|=1} \left\{\langle A, vv^T\rangle\right\}.$$

Motivated by the above second equality, we relax the set $\left\{v \in \mathbb{R}^n \middle| \|v\| = 1\right\}$ to the one $\mathcal{W} = \left\{W \succeq 0 \middle| tr(W) = 1\right\}$, it is not difficult to see that

$$\lambda_1(A) = \max\left\{\langle A, W\rangle \middle| W \in \mathcal{W}\right\}. \tag{2.3}$$

Similarly, it holds $a\lambda_1(A) = \max\left\{\langle A, W\rangle \middle| W \in a\mathcal{W}\right\}$.

The next result characterizes a key property of $\partial\lambda_1(A)$, which plays a crucial role on formulating an approximate of $a\mathcal{W}$ in a more tractable way, see (3.1).

Proposition 2.2. $\partial\lambda_1(A) = \left\{PVP^T \middle| tr(V) = 1, V \succeq 0\right\}$, *where the columns of* P *form an orthonormal basis of the eigenspace to the maximum eigenvalue of* A.

Proof. This result was already established in the light of its geometrical descriptions in [8, 38, 39]. □

In many practical problems, exact eigenvalues and eigenvectors are unavailable, which mostly results from truncating data for storage or finite-precision arithmetic. Therefore it is more reasonable to consider the approximate subdifferential of the maximum eigenvalue function for a symmetric matrix A, which is so called ϵ-subdifferential and visualized by the following proposition.

Proposition 2.3. *[17] For* $\epsilon \geq 0$, *it holds* $\partial_\epsilon\lambda_1(A) = \left\{W \in \mathcal{W} \middle| \langle W, A\rangle \geq \lambda_1(A) - \epsilon\right\}$.

Using again a chain rule [18, Theorem XI.3.2.1], we have

$$\partial_\epsilon f(X; y) = \left\{(\mathcal{Q}(X - W); b - \mathcal{A}(W)) \middle| W \in a\partial_\epsilon\lambda_1(-C - \mathcal{A}^*(y) - \mathcal{Q}(X))\right\}. \tag{2.4}$$

For later use, we adopt the *Ritz triplet* introduced in [34]: $(\lambda_{ap}, v_{ap}, \epsilon) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}$ is a *Ritz triplet* for $\lambda_1(A)$ if

$$\lambda_{ap} = v_{ap}^T A v_{ap}, \ \|v_{ap}\| = 1, \ \epsilon = \|Av_{ap} - \lambda_{ap}v_{ap}\|, \tag{2.5}$$

$$|\lambda_i(A) - \lambda_{ap}| \geq |\lambda_1(A) - \lambda_{ap}| \text{ for } i = 2, \cdots, n. \tag{2.6}$$

We will see that a *Ritz triplet* depicts a pair of approximation eigenvalue and eigenvector for a symmetric matrix.

Proposition 2.4. *If* $(\lambda_{ap}, v_{ap}, \epsilon)$ *is a* Ritz triplet *for* $\lambda_1(A)$, *then*

$$\lambda_{ap} \leq \lambda_1 \leq \lambda_{ap} + \epsilon \textit{ and } v_{ap}v_{ap}^T \in \partial_\epsilon\lambda_1(A).$$

Proof. Suppose that A has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and corresponding orthonormal eigenvectors $z_1, z_2, \cdots, z_n$, then v_{ap} has the expression $v_{ap} = \sum_{i=1}^n d_i z_i$ with $\sum_{i=1}^n d_i^2 = 1$. From (2.3) and the first two equalities in (2.5), we get $\lambda_{ap} \leq \lambda_1$.

On the other hand,

$$\epsilon = \left\|\sum_{i=1}^n \lambda_i d_i z_i - \sum_{i=1}^n \lambda_{ap} d_i z_i\right\| = \sqrt{\sum_{i=1}^n (\lambda_i - \lambda_{ap})^2 d_i^2} \geq \sqrt{\sum_{i=1}^n (\lambda_1 - \lambda_{ap})^2 d_i^2} = |\lambda_1 - \lambda_{ap}| = \lambda_1 - \lambda_{ap}.$$

where the inequality follows from (2.6). Therefore $\lambda_{ap} \leq \lambda_1 \leq \lambda_{ap} + \epsilon$.

By Proposition 2.3, it is easy to see that $v_{ap}v_{ap}^T \in \partial_\epsilon\lambda_1(A)$. □

In what follows, corresponding to a *Ritz triplet* $(\lambda_{ap}, v_{ap}, \epsilon)$ for $\lambda_1(-C - \mathcal{A}^*(y) - \mathcal{Q}(X))$, we define the inexact computation of $f(X; y)$ by

$$f_l(X; y) = a\lambda_{ap} + b^T y + \langle X, \mathcal{Q}(X)\rangle/2 \text{ and } f_u(X; y) = f_l(X; y) + a\epsilon.$$

Obviously, $f_l(X; y) \leq f(X; y) \leq f_u(X; y)$ for any $(X; y) \in \mathcal{S}^n \times \mathbb{R}^m$.

3. The Algorithm and Error bound

In this section, we focus on presenting the inexact spectral bundle method for solving (Eigform) in the inexact setting and providing its convergence and error-bound analysis. Although the basic ideas of the proofs of convergence theorems are similar to those presented in [14], their details are more involved due to the existence of quadratic term in the objective function of the problem. Furthermore, we provide a global error bound under mild assumptions. We introduce firstly the main ingredients of this method.

3.1. **The approximation of f.** Adopting the ideas from [14] and [38], we use

$$\widehat{\mathcal{W}}^k = \left\{ P_k V P_k^T + \xi \overline{W}_k \middle| tr(V) + \xi = a, V \in \mathcal{S}_+^{r_k}, \xi \geq 0 \right\}, \tag{3.1}$$

as an approximation of $a\mathcal{W}$ with $\widehat{\mathcal{W}}^k \subseteq a\mathcal{W}$, where $\overline{W}_k \in \mathcal{W}$ is the so call *aggregate subgradient*, and $P_k \in \mathbb{R}^{n \times r_k}$ is an orthonormal matrix, $r_k \geq 1$. A special advantage of this approximation is that it can keep the number r_k of columns of P_k small. The idea behind (3.1) is explained in [38].

For any given $W \in a\mathcal{W}$, let the *approximate partial linearization* of f be

$$\begin{aligned} f_W(X; y) &= \langle -C - \mathcal{A}^*(y) - \mathcal{Q}(X), W \rangle + b^T y + \langle X, \mathcal{Q}(X) \rangle / 2 \\ &= \langle -C - \mathcal{A}^*(y) - \mathcal{Q}(X), W \rangle + b^T y + \langle \widehat{X}^k, \mathcal{Q}(\widehat{X}^k) \rangle / 2 \\ &+ \langle X - \widehat{X}^k, \mathcal{Q}(\widehat{X}^k) \rangle + \langle X - \widehat{X}^k, \mathcal{Q}(X - \widehat{X}^k) \rangle / 2 \end{aligned} \tag{3.2}$$

and define an approximation of $f(X; y)$ as $f_{\widehat{\mathcal{W}}^k}(X; y) = \max_{W \in \widehat{\mathcal{W}}^k} f_W(X; y)$, which is different from the *cutting plane* model of f

$$\max_{i=1,\cdots,k} \left\{ \bar{f}_i(X; y) := f(X^i; y^i) + \langle \mathcal{Q}(X^i - p_i p_i^T), X - X^i \rangle + \langle b - \mathcal{A}(p_i p_i^T), y - y^i \rangle \right\},$$

where p_i is a normalized eigenvector to $\lambda_1(-C - \mathcal{A}^*(y^i) - \mathcal{Q}(X^i))$ for $i = 1, \cdots, k$. In effect, by setting $\widehat{\mathcal{W}}^k = \operatorname{conv}\left\{ p_i p_i^T \middle| i = 1, \cdots, k \right\}$, we have

$$f_{\widehat{\mathcal{W}}^k}(X; y) = \max_{i=1,\cdots,k} \left\{ \bar{f}_i(X; y) + \langle X - X^i, \mathcal{Q}(X - X^i) \rangle / 2 \right\}.$$

This difference is not likely to happen for the linear semidefinite programming.

Remark 3.1. $f_{\widehat{\mathcal{W}}^k}(X; y)$ does minorize $f(X; y)$ for any $(X; y) \in \mathcal{S}^n \times \mathbb{R}^m$. □

In order to determine a new *trial point* $(X^{k+1}; y^{k+1})$ from the current *stability center* $(\widehat{X}^k; \hat{y}^k)$, we use the model stabilized by a quadratic penalty or regularization. For each k, we define

$$L^k(X, y, W) = f_W(X; y) + u_k (\|X - \widehat{X}^k\|^2_{\mathcal{M}_{X,k}} + \|y - \hat{y}^k\|^2_{\mathbf{M}_{y,k}})/2. \tag{3.3}$$

Recall that u_k, $\mathcal{M}_{X,k}$ and $\mathbf{M}_{y,k}$ have been mentioned in (1.2).

Now we can define a sequence of minimization subproblems $\min_{X,y} \phi^k(X; y)$ incorporating the so called *Moreau-Yosida regularization*, where

$$\phi^k(X; y) = \max_{W \in \widehat{\mathcal{W}}^k} L^k(X, y, W) = f_{\widehat{\mathcal{W}}^k}(X; y) + u_k (\|X - \widehat{X}^k\|^2_{\mathcal{M}_{X,k}} + \|y - \hat{y}^k\|^2_{\mathbf{M}_{y,k}})/2. \tag{3.4}$$

In order to figure out the properties of $\phi^k(X; y)$, we need the following two propositions.

Proposition 3.1. *We can rewrite $f_{\widehat{\mathcal{W}}^k}(X; y)$ as*

$$a \max \left\{ \lambda_1 [P_k^T (-C - \mathcal{A}^*(y) - \mathcal{Q}(X)) P_k], \langle -C - \mathcal{A}^*(y) - \mathcal{Q}(X), \overline{W}_k \rangle \right\} + b^T y + \langle X, \mathcal{Q}(X) \rangle / 2.$$

Proof. For any fixed $(X;y)$, thanks to the expression of $f_{\widehat{\mathcal{W}}^k}(X;y)$, we see that it is equal to the maximum over the extreme points of $\widehat{\mathcal{W}}^k$.

Since $a\overline{W}_k \in \widehat{\mathcal{W}}^k$ and $\left\{P_kUP_k^T \middle| \mathrm{tr}(U)=a, U\in\mathcal{S}_+^{r_k}\right\} =: \mathcal{W}_0 \subseteq \widehat{\mathcal{W}}^k$, it holds $\mathrm{conv}(\{a\overline{W}_k\}\cup \mathcal{W}_0) \subseteq \widehat{\mathcal{W}}^k$. Conversely, for any $W = P_kVP_k^T + \xi\overline{W}_k \in \widehat{\mathcal{W}}^k$ with $0<\xi<a$, set $U = \frac{a}{a-\xi}V$, then $\mathrm{tr}(U) = \frac{a}{a-\xi}\mathrm{tr}(V) = a$. Thus $\mathrm{conv}(\{a\overline{W}_k\}\cup\mathcal{W}_0) = \widehat{\mathcal{W}}^k$. Therefore, $f_{\widehat{\mathcal{W}}^k}(X;y) = \max\left\{f_{a\overline{W}_k}(X;y), f_{\mathcal{W}_0}(X;y)\right\}$.

On the other hand,

$$\begin{aligned} f_{\mathcal{W}_0}(X;y) =& \max_{U\in\left\{U\in\mathcal{S}_+^{r_k}\middle|\mathrm{tr}(U)=a\right\}} \left\{\langle P_k^T(-C-\mathcal{A}^*(y)-\mathcal{Q}(X))P_k, U\rangle\right\} + b^Ty + \langle X,\mathcal{Q}(X)\rangle/2 \\ =& a\lambda_1[P_k^T(-C-\mathcal{A}^*(y)-\mathcal{Q}(X))P_k] + b^Ty + \langle X,\mathcal{Q}(X)\rangle/2, \end{aligned}$$

the proof is completed. □

For later use, for any k, we let $\mathcal{T}_k$ be $\mathcal{Q} + u_k\mathcal{M}_{X,k}$, which is a self-adjoint positive definite linear operator acting on $\mathcal{S}^n$.

The following proposition handles how to compute a trial point at each iteration of the algorithm. In essence, we obtain $(X^{k+1};y^{k+1})$ via solving the minimization subproblem $\min_{X,y}\phi^k(X;y)$.

Proposition 3.2. *The minimizer of $\phi^k(X;y)$ over $\mathcal{S}^n\times\mathbb{R}^m$ possesses the formulated below property:*

$$\min_{X,y}\max_{W\in\widehat{\mathcal{W}}^k} L^k(X,y,W) = L^k(X^{k+1},y^{k+1},W^{k+1}) = \max_{W\in\widehat{\mathcal{W}}^k}\min_{X,y} L^k(X,y,W),$$

where

$$X^{k+1} = \widehat{X}^k + \mathcal{T}_k^{-1}(\mathcal{Q}(W^{k+1}-\widehat{X}^k)),\ y^{k+1} = \hat{y}^k + \frac{1}{u_k}\mathbf{M}_{y,k}^{-1}(\mathcal{A}(W^{k+1})-b), \tag{3.5}$$

uniquely, and W^{k+1} is an optimal solution of

$$\text{(SQSDP)}\qquad \begin{array}{ll} \min & \frac{1}{2}\left\|\mathcal{Q}(W-\widehat{X}^k)\right\|^2_{\mathcal{T}_k^{-1}} + \frac{1}{2u_k}\left\|b-\mathcal{A}(W)\right\|^2_{\mathbf{M}_{y,k}^{-1}} - b^T\hat{y}^k \\ & +\langle C+\mathcal{A}^*(\hat{y}^k)+\mathcal{Q}(\widehat{X}^k), W\rangle - \frac{1}{2}\langle\widehat{X}^k,\mathcal{Q}(\widehat{X}^k)\rangle \\ \text{s.t.} & W = P_kVP_k^T + \xi\overline{W}_k \\ & tr(V)+\xi = a \\ & V\succeq 0, \xi\geq 0. \end{array}$$

Proof. By virtue of Proposition 3.1, $f_{\widehat{\mathcal{W}}^k}(X;y)$ is equivalent to

$$\begin{array}{ll} \min & a\lambda + b^Ty + \frac{1}{2}\langle X,\mathcal{Q}(X)\rangle \\ \text{s.t.} & P_k^T(-C-\mathcal{A}^*(y)-\mathcal{Q}(X))P_k \preceq \lambda I \\ & \langle -C-\mathcal{A}^*(y)-\mathcal{Q}(X), \overline{W}_k\rangle \leq \lambda. \end{array}$$

Thus, from (3.4), $\min_{X,y}\phi^k(X;y)$ can be reformulated as

$$\begin{array}{ll} \min & a\lambda + b^Ty + \frac{1}{2}\langle X,\mathcal{Q}(X)\rangle + \frac{u_k}{2}(\|X-\widehat{X}^k\|^2_{\mathcal{M}_{X,k}} + \|y-\hat{y}^k\|^2_{\mathbf{M}_{y,k}}) \\ \text{s.t.} & \lambda I + P_k^T(\mathcal{A}^*(y)+\mathcal{Q}(X)+C)P_k \succeq 0 \\ & \lambda + \langle\mathcal{A}^*(y)+\mathcal{Q}(X)+C,\overline{W}_k\rangle \geq 0. \end{array}$$

We define the Langrangian of the above problem by

$$\begin{aligned} L(X,y,\lambda,V,\xi) =& a\lambda + b^Ty + \frac{1}{2}\langle X,\mathcal{Q}(X)\rangle + \frac{u_k}{2}(\|X-\widehat{X}^k\|^2_{\mathcal{M}_{X,k}} + \|y-\hat{y}^k\|^2_{\mathbf{M}_{y,k}}) \\ & - \langle V, \lambda I + P_k^T(\mathcal{A}^*(y)+\mathcal{Q}(X)+C)P_k\rangle - \xi(\lambda + \langle\mathcal{A}^*(y)+\mathcal{Q}(X)+C,\overline{W}_k\rangle) \end{aligned}$$

By the Lagrange Duality theorem, a direct computation implies that

$$\begin{aligned}
\frac{\partial L}{\partial X} &= \mathcal{Q}(X) - \mathcal{Q}(P_k V P_k^T + \xi \overline{W}_k) + u_k \mathcal{M}_{X,k}(X - \widehat{X}^k) &&= 0,\\
\frac{\partial L}{\partial y} &= b - \mathcal{A}(P_k V P_k^T + \xi \overline{W}_k) + u_k \mathbf{M}_{y,k}(y - \hat{y}^k) &&= 0,\\
\frac{\partial L}{\partial \lambda} &= a - \langle V, I\rangle - \xi &&= 0.
\end{aligned}$$

Set $W = P_k V P_k^T + \xi \overline{W}_k$, it follows

$$X = \widehat{X}^k + \mathcal{T}_k^{-1}(\mathcal{Q}(W - \widehat{X}^k)) =: X^k_{min}(W), \tag{3.6}$$

$$y = \hat{y}^k + \frac{1}{u_k}\mathbf{M}_{y,k}^{-1}(\mathcal{A}(W) - b) =: y^k_{min}(W), \tag{3.7}$$

Moreover, such $(X;y)$ is unique since $\phi^k(X;y)$ is strictly convex. As for the solution (V^*, ξ^*) of the above linear systems, let $W^{k+1} = P_k V^* P_k^T + \xi^* \overline{W}_k$, we get the corresponding $(X^{k+1}; y^{k+1})$ with the desired expressions.

On the other hand, together with (3.3), we obtain

$$\min_{X,y} L^k(X,y,W) = L^k(X^k_{min}(W), y^k_{min}(W), W) = -\frac{1}{2}\|\mathcal{Q}(W - \widehat{X}^k)\|^2_{\mathcal{T}_k^{-1}} - \frac{1}{2u_k}\|b - \mathcal{A}(W)\|^2_{\mathbf{M}_{y,k}^{-1}}$$
$$+ b^T\hat{y}^k - \langle C + \mathcal{A}^*(\hat{y}^k) + \mathcal{Q}(\widehat{X}^k), W\rangle + \frac{1}{2}\langle \widehat{X}^k, \mathcal{Q}(\widehat{X}^k)\rangle.$$

By convex minimax duality, $\min_{X,y} \phi^k(X;y) = \min_{X,y} \max_{W\in\widehat{\mathcal{W}}^k} L^k(X,y,W) = \max_{W\in\widehat{\mathcal{W}}^k} \min_{X,y} L^k(X,y,W)$. Switching the optimization direction in $\max_{W\in\widehat{\mathcal{W}}^k}$ and the sign in $\min_{X,y} L^k(X,y,W)$, we get (SQSDP) . The proof is completed. □

In the light of Proposition 3.2,

$$(X^{k+1}; y^{k+1}) = arg\min_{X,y}\left\{ f_{\widehat{\mathcal{W}}^k}(X;y) + \frac{u_k}{2}(\|X - \widehat{X}^k\|^2_{\mathcal{M}_{X,k}} + \|y - \hat{y}^k\|^2_{\mathbf{M}_{y,k}})\right\},$$

the necessary and sufficient optimality conditions for the strictly convex problem imply

$$0 \in \partial f_{\widehat{\mathcal{W}}^k}(X^{k+1}; y^{k+1}) + u_k(\mathcal{M}_{X,k}(X^{k+1} - \widehat{X}^k); \mathbf{M}_{y,k}(y^{k+1} - \hat{y}^k)) \Leftrightarrow \exists\ (g^k_X; g^k_y) \in \partial f_{\widehat{\mathcal{W}}^k}(X^{k+1}; y^{k+1}) \text{ s.t.}$$

$$(g^k_X; g^k_y) + u_k(\mathcal{M}_{X,k}(X^{k+1} - \widehat{X}^k); \mathbf{M}_{y,k}(y^{k+1} - \hat{y}^k)) = 0. \tag{3.8}$$

Comparing (3.8) with the expressions of $(X^{k+1}; y^{k+1})$, we obtain

$$(g^k_X; g^k_y) = (\mathcal{Q}(X^{k+1} - W^{k+1}); b - \mathcal{A}(W^{k+1})) \in \partial f_{\widehat{\mathcal{W}}^k}(X^{k+1}; y^{k+1}). \tag{3.9}$$

At this point, we need the linearization of $f_{\widehat{\mathcal{W}}^k}$ for brevity. With the help of the above subgradient, we define the linearization of $f_{\widehat{\mathcal{W}}^k}$ at $(X^{k+1}; y^{k+1})$ as

$$\bar{f}^k_{\mathcal{W}}(X;y) = f_{W^{k+1}}(X^{k+1}; y^{k+1}) + \langle g^k_X, X - X^{k+1}\rangle + \langle g^k_y, y - y^{k+1}\rangle. \tag{3.10}$$

In addition,

$$(g^k_X; g^k_y) \in \partial_{\varepsilon_k} f(\widehat{X}^k; \hat{y}^k), \tag{3.11}$$

where

$$0 \le \varepsilon_k := f(\widehat{X}^k; \hat{y}^k) - \bar{f}^k_{\mathcal{W}}(\widehat{X}^k; \hat{y}^k). \tag{3.12}$$

To see this, note that in view of (3.9) and Proposition 3.2, for any $(X;y) \in \mathcal{S}^n \times \mathbb{R}^m$,

$$f_{\widehat{\mathcal{W}}^k}(X;y) \ge f_{\widehat{\mathcal{W}}^k}(X^{k+1}; y^{k+1}) + \langle g^k_X, X - X^{k+1}\rangle + \langle g^k_y, y - y^{k+1}\rangle = \bar{f}^k_{\mathcal{W}}(X;y),$$

then $\varepsilon_k \geq f_{\widehat{\mathcal{W}}^k}(\widehat{X}^k;\hat{y}^k) - \bar{f}^k_{\mathcal{W}}(\widehat{X}^k;\hat{y}^k) \geq 0$.

Hence

$$\begin{aligned}&f(\widehat{X}^k;\hat{y}^k) + \langle X - \widehat{X}^k, \mathcal{Q}(X^{k+1} - W^{k+1})\rangle + \langle y - \hat{y}^k, b - \mathcal{A}(W^{k+1})\rangle - \varepsilon_k \\ =&f_{\widehat{\mathcal{W}}^k}(X^{k+1};y^{k+1}) + \langle X - X^{k+1}, \mathcal{Q}(X^{k+1} - W^{k+1})\rangle + \langle y - y^{k+1}, b - \mathcal{A}(W^{k+1})\rangle \\ \leq&f_{\widehat{\mathcal{W}}^k}(X;y) \leq f(X;y).\end{aligned}$$

The approximate subgradient relation in (3.11) can certainly be employed to derive an optimality estimate. If doing so, owing to (3.12), one needs to compute $f(\widehat{X}^k;\hat{y}^k)$ exactly, which may be expensive. Thus we would like to provide a relaxation of ε_k in the following corollary, which enables us to make an optimality estimate tractable.

Corollary 3.1. *Suppose that* $(\lambda^{k+1}_{ap}, v^{k+1}_{ap}, \epsilon^{k+1})$ *and* $(\hat{\lambda}^k_{ap}, \hat{v}^k_{ap}, \hat{\epsilon}^k)$ *are Ritz triplets for* $\lambda_1(-C - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(X^{k+1}))$ *and* $\lambda_1(-C - \mathcal{A}^*(\hat{y}^k) - \mathcal{Q}(\widehat{X}^k))$, *respectively, we have*

$$(g^k_X;g^k_y) \in \partial_{\eta^k} f(X^{k+1};y^{k+1}) \text{ and } (g^k_X;g^k_y) \in \partial_{\tau^k} f(\widehat{X}^k;\hat{y}^k) \tag{3.13}$$

where

$$\eta^k = a(\lambda^{k+1}_{ap} + \epsilon^{k+1}) - \langle W^{k+1}, -C - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(X^{k+1})\rangle, \tag{3.14}$$

$$\tau^k = f_l(\widehat{X}^k;\hat{y}^k) + a\hat{\epsilon}^k - \bar{f}^k_{\mathcal{W}}(\widehat{X}^k;\hat{y}^k). \tag{3.15}$$

Proof. In view of (3.9) and the definition of the Ritz triplet for $\lambda_1(-C - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(X^{k+1}))$, for any $(X;y) \in \mathcal{S}^n \times \mathbb{R}^m$, it holds

$$\begin{aligned}f(X;y) \geq& f_{\widehat{\mathcal{W}}^k}(X;y) \geq f_{\widehat{\mathcal{W}}^k}(X^{k+1};y^{k+1}) + \langle g^k_X, X - X^{k+1}\rangle + \langle g^k_y, y - y^{k+1}\rangle \\ \geq& f(X^{k+1};y^{k+1}) + \langle g^k_X, X - X^{k+1}\rangle + \langle g^k_y, y - y^{k+1}\rangle \\ &- (a\lambda^{k+1}_{ap} + a\epsilon^{k+1} - \langle W^{k+1}, -C - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(X^{k+1})\rangle).\end{aligned}$$

Note that $W^{k+1} \in a\mathcal{W}$, it follows

$$\langle W^{k+1}, -C - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(X^{k+1})\rangle \leq a\lambda_1(-C - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(X^{k+1})),$$

whence $\eta^k \geq 0$ and $(g^k_X;g^k_y) \in \partial_{\eta^k} f(X^{k+1};y^{k+1})$.

Putting together (3.12) and the definition of the Ritz triplet for $\lambda_1(-C - \mathcal{A}^*(\hat{y}^k) - \mathcal{Q}(\widehat{X}^k))$, we obtain $\tau^k \geq \varepsilon_k \geq 0$, whence it satisfies $(g^k_X;g^k_y) \in \partial_{\tau^k} f(\widehat{X}^k;\hat{y}^k)$ by (3.11). □

Set

$$d^k = (d^k_X;d^k_y) = (X^{k+1} - \widehat{X}^k; y^{k+1} - \hat{y}^k),\ \ l^k_d = \|d^k_X\|^2_{\mathcal{M}_{X,k}} + \|d^k_y\|^2_{\mathbf{M}_{y,k}}. \tag{3.16}$$

In accordance with (3.8), we have, for $i = 1, 2$,

$$\begin{aligned}&\langle g^k_X, X - X^i\rangle + \langle g^k_y, y - y^i\rangle \\ =&u_k(\langle \mathcal{M}^{1/2}_{X,k}(\widehat{X}^k - X^{k+1}), \mathcal{M}^{1/2}_{X,k}(X - X^i)\rangle + \langle \mathbf{M}^{1/2}_{y,k}(\hat{y}^k - y^{k+1}), \mathbf{M}^{1/2}_{y,k}(y - y^i)\rangle) \\ \leq&c_k(\|u_k\mathcal{M}^{1/2}_{X,k}(d^k_X)\|^2 + \|u_k\mathbf{M}^{1/2}_{y,k}d^k_y\|^2)^{1/2}(\|X - X^i\|^2 + \|y - y^i\|^2)^{1/2} \\ =&c_k u_k\sqrt{l^k_d}(\|X - X^i\|^2 + \|y - y^i\|^2)^{1/2}\end{aligned}$$

where $c_k = \max\{\|\mathcal{M}^{1/2}_{X,k}\|, \|\mathbf{M}^{1/2}_{y,k}\|\} > 0$, $(X^1;y^1) = (\widehat{X}^k;\hat{y}^k)$ and $(X^2;y^2) = (X^{k+1};y^{k+1})$.

We now have the necessary ingredients for deriving the optimality estimate. The approximate subgradients in (3.13) yield the optimality estimate

$$f(\widehat{X}^k;\hat{y}^k) \le f(X;y) + c_k u_k \sqrt{l_d^k}(\|X-\widehat{X}^k\|^2 + \|y-\hat{y}^k\|^2)^{1/2} + \tau^k \tag{3.17}$$

$$f(X^{k+1};y^{k+1}) \le f(X;y) + c_k u_k \sqrt{l_d^k}(\|X-X^{k+1}\|^2 + \|y-y^{k+1}\|^2)^{1/2} + \eta^k \tag{3.18}$$

for all $(X;y) \in \mathcal{S}^n \times \mathbb{R}^m$.

The inequalities (3.17) and (3.18) say that the point $(\widehat{X}^k;\hat{y}^k)$ or $(X^{k+1};y^{k+1})$ is optimal if the optimality measure

$$\nu^k = \max\left\{c_k u_k \sqrt{l_d^k},\ \min\{\eta^k,\tau^k\}\right\} \tag{3.19}$$

is zero, which will be established in the following proposition providing a stopping criterion for our algorithm. Meanwhile, the coming result also tackles the case of a finite number of iterations for our method.

Proposition 3.3. *For $W \in \widehat{\mathcal{W}}^k$, let $X^k_{min}(W)$ and $y^k_{min}(W)$ be defined by (3.6) and (3.7) respectively, it holds*

$$L^k(X,y,W) = L^k(X^k_{min}(W), y^k_{min}(W), W) + \frac{1}{2}\|X - X^k_{min}(W)\|^2_{\mathcal{T}_k} + \frac{u_k}{2}\|y - y^k_{min}(W)\|^2_{\mathbf{M}_{y,k}}. \tag{3.20}$$

Moreover, suppose that $(X^{k+1};y^{k+1})$ satisfies (3.5) and W^{k+1} is an optimal solution of (SQSDP), then

$$L^k(X^{k+1},y^{k+1},W^{k+1}) \le f(\widehat{X}^k;\hat{y}^k), \tag{3.21}$$

and if $\nu^k = 0$ then $f(\widehat{X}^k;\hat{y}^k) = f_{W^{k+1}}(X^{k+1};y^{k+1})$ and $(\widehat{X}^k;\hat{y}^k)$ is optimal.

Proof. By (3.3), together with (3.6) and (3.7), it follows

$$\begin{aligned}
& L^k(X,y,W) - L^k(X^k_{min}(W), y^k_{min}(W), W) \\
=& \langle \mathcal{A}(W) - b, y^k_{min}(W) - y\rangle + \langle \mathcal{Q}(W - \widehat{X}^k), X^k_{min}(W) - X\rangle \\
& + \frac{1}{2}(\|X\|^2_{\mathcal{T}_k} - \|X^k_{min}(W)\|^2_{\mathcal{T}_k} - 2\langle X - X^k_{min}(W), \mathcal{T}_k(\widehat{X}^k)\rangle) \\
& + \frac{u_k}{2}(\|y\|^2_{\mathbf{M}_{y,k}} - \|y^k_{min}(W)\|^2_{\mathbf{M}_{y,k}} - 2\langle y - y^k_{min}(W), \mathbf{M}_{y,k}(\hat{y}^k)\rangle) \\
=& u_k\langle y^k_{min}(W), \mathbf{M}_{y,k}(y^k_{min}(W) - y)\rangle + \langle X^k_{min}(W), \mathcal{T}_k(X^k_{min}(W) - X)\rangle \\
& + \frac{1}{2}(\|X\|^2_{\mathcal{T}_k} - \|X^k_{min}(W)\|^2_{\mathcal{T}_k} - 2\langle X - X^k_{min}(W), \mathcal{T}_k(\widehat{X}^k)\rangle) \\
& + \frac{u_k}{2}(\|y\|^2_{\mathbf{M}_{y,k}} - \|y^k_{min}(W)\|^2_{\mathbf{M}_{y,k}} - 2\langle y - y^k_{min}(W), \mathbf{M}_{y,k}(\hat{y}^k)\rangle) \\
=& \frac{1}{2}\|X - X^k_{min}(W)\|^2_{\mathcal{T}_k} + \frac{u_k}{2}\|y - y^k_{min}(W)\|^2_{\mathbf{M}_{y,k}}.
\end{aligned}$$

From Proposition 3.2, it yields

$$\begin{aligned}
& L^k(X^{k+1}, y^{k+1}, W^{k+1}) \le L^k(\widehat{X}^k, \hat{y}^k, W^{k+1}) \\
& \le \max_{W \in a\mathcal{W}} \langle -C - \mathcal{A}^*(\hat{y}^k) - \mathcal{Q}(\widehat{X}^k), W\rangle + b^T\hat{y}^k + \frac{1}{2}\langle \widehat{X}^k, \mathcal{Q}(\widehat{X}^k)\rangle = f(\widehat{X}^k;\hat{y}^k).
\end{aligned}$$

If $\nu^k = 0$, then $l_d^k = 0$ and $\min\{\eta^k, \tau^k\} = 0$.

Since both $\mathcal{M}_{X,k}$ and $\mathbf{M}_{y,k}$ are invertible, we get $X^{k+1} = \widehat{X}^k$ and $y^{k+1} = \hat{y}^k$. Furthermore, combing with (3.8), we have $(g_X^k; g_y^k) = (0; 0)$. Therefore, $(\mathcal{Q}(X^{k+1} - W^{k+1}); b - \mathcal{A}(W^{k+1})) \in \partial f(\widehat{X}^k; \hat{y}^k)$ because at least one of η^k and τ^k is equal to zero. If $\eta^k = 0$, then $a\lambda_1(-C - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(X^{k+1})) \le \langle W^{k+1}, -C - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(X^{k+1})\rangle$. Consequently, $f_{W^{k+1}}(X^{k+1}; y^{k+1}) = f(X^{k+1}; y^{k+1}) = f(\widehat{X}^k; \hat{y}^k)$. If $\tau^k = 0$, then $f(\widehat{X}^k; \hat{y}^k) \le f_{W^{k+1}}(X^{k+1}; y^{k+1})$, hence $f(\widehat{X}^k; \hat{y}^k) = f_{W^{k+1}}(X^{k+1}; y^{k+1})$.

According to (3.6) and (3.7), $\mathcal{Q}(W^{k+1} - X^{k+1}) = 0$ and $b - \mathcal{A}(W^{k+1}) = 0$. Then $0 \in \partial f(X^{k+1}; y^{k+1}) = \partial f(\widehat{X}^k; \hat{y}^k)$. Hence $(\widehat{X}^k; \hat{y}^k)$ is optimal. □

3.2. **Updating $\widehat{\mathcal{W}}^k$.** The updating rule for $\widehat{\mathcal{W}}^k$ is identical to the one presented in [14], we review formulas that are necessary for the purpose of this paper.

Let (V^*, ξ^*) denote the optimal solution for (SQSDP) that gives rise to W^{k+1}, and $R\Lambda R^T$ denote an eigenvalue decomposition of V^* with $RR^T = I$ and $\Lambda = Diag(\lambda_1, \cdots, \lambda_{r_k})$, $\lambda_1 \ge \cdots \ge \lambda_{r_k}$. We break up R into two parts, $R = [R_1, R_2]$, R_1 containing the eigenvectors to the "large" eigenvalues of V^*. Correspondingly, we partition Λ into two smaller diagonal matrices Λ_1 and Λ_2.

Via the scheme of the proximal bundle method, at each iteration k, let v_{ap}^{k+1} be a normalized approximate eigenvector corresponding to the approximate maximum eigenvalue of $-C - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(X^{k+1})$, we should add new ϵ-subgradient information $W_{ap}^{k+1} = av_{ap}^{k+1}(v_{ap}^{k+1})^T \in a\mathcal{W}$ to $\widehat{\mathcal{W}}^k$ by adding the new v_{ap}^{k+1} as orthonormalized column to P for increasing the number of columns. In order to update $\widehat{\mathcal{W}}^k$, we can update P_k and $\overline{W}_k$ by

$$P_{k+1} = orth[P_k R_1, v_{ap}^{k+1}], \tag{3.22}$$

$$\overline{W}_{k+1} = \frac{(P_k R_2)\Lambda_2 (P_k R_2)^T + \xi^* \overline{W}_k}{\mathrm{tr}(\Lambda_2) + \xi^*}. \tag{3.23}$$

We will see from the convergence analysis that

$$W^{k+1},\ W_{ap}^{k+1} \in \widehat{\mathcal{W}}^{k+1} \tag{3.24}$$

is all that is needed to guarantee convergence. The following proposition (see [14]) shows that update formulas (3.22) and (3.23) have the desired properties.

Proposition 3.4. *For $W_{ap}^{k+1} = av_{ap}^{k+1}(v_{ap}^{k+1})^T$, update formulas (3.22) and (3.23) ensure that P_{k+1} is orthnormal, $\overline{W}_{k+1} \in \mathcal{W}$, and (3.24) is satisfied for $\widehat{\mathcal{W}}^{k+1}$ of the form (3.1).*

3.3. **Updating weight u_k.** The choice of weight u_k is somewhat of an art, which impacts significantly both global convergence in theory and efficiency in practice, for instance, as indicated in [16], a suitable adjustment of u_k shall guide $f_{\widehat{\mathcal{W}}^k}$ to the area where it is a reliable approximation of f. There are several wise update rules published in the literature; see, for instance, [5, 24, 44]. The idea of the updating rule here is the same as that in [24], we outline the main results here.

First we detect whether u_k is too large.

Let $d = (d_X; d_y) = (X - \widehat{X}^k; y - \hat{y}^k)$, it will be convenient to have an alternate description of $f_W(X; y)$, which is in terms of d. With the help of (3.2), it yields

$$f_W(X; y) = f_W(\widehat{X}^k; \hat{y}^k) + \langle b - \mathcal{A}(W), d_y\rangle + \langle \mathcal{Q}(\widehat{X}^k - W), d_X\rangle + \langle d_X, \mathcal{Q}(d_X)\rangle/2 =: f_W^{dr}((\widehat{X}^k; \hat{y}^k), d).$$

Correspondingly, $f_{\widehat{\mathcal{W}}^k}^{dr}((\widehat{X}^k; \hat{y}^k), d) = \max\limits_{W \in \widehat{\mathcal{W}}^k} f_W^{dr}((\widehat{X}^k; \hat{y}^k), d)$, thus $f_{\widehat{\mathcal{W}}^k}^{dr}((\widehat{X}^k; \hat{y}^k), 0) = f_{\widehat{\mathcal{W}}^k}(\widehat{X}^k; \hat{y}^k) \le f(\widehat{X}^k; \hat{y}^k)$. By virtue of the convexity of $f_{\widehat{\mathcal{W}}^k}^{dr}((\widehat{X}^k; \hat{y}^k), d)$ with reference to d and $f_l(\widehat{X}^k; \hat{y}^k) \le$

$f(\widehat{X}^k;\hat{y}^k)$, it comes up for $\kappa \in [0,1]$

$$f^{dr}_{\widehat{\mathcal{W}}^k}((\widehat{X}^k;\hat{y}^k),\kappa d^k) \leq (1-\kappa) f^{dr}_{\widehat{\mathcal{W}}^k}((\widehat{X}^k;\hat{y}^k),0) + \kappa f^{dr}_{\widehat{\mathcal{W}}^k}((\widehat{X}^k;\hat{y}^k),d^k) \leq f(\widehat{X}^k;\hat{y}^k) - \kappa\vartheta^k,$$

with

$$\vartheta^k = f_l(\widehat{X}^k;\hat{y}^k) - f_{W^{k+1}}(X^{k+1};y^{k+1}). \tag{3.25}$$

Therefore ϑ^k estimates the descent obtained from our model and will subsequently serve as *predicted descent* of f. We emphasis that ϑ^k may be nonpositive and $f_l(\widehat{X}^k;\hat{y}^k) - f_u(X^{k+1};y^{k+1}) \geq \kappa\vartheta^k$ can guarantee $\vartheta^k > 0$ for $\kappa \in (0,1)$.

We now are in a position to derive the reduction of u_k.

It may be reduced if $f_{\widehat{\mathcal{W}}^k}$ is close to f_u at $(X^{k+1};y^{k+1})$, which is measured by

$$f_l(\widehat{X}^k;\hat{y}^k) - f_u(X^{k+1};y^{k+1}) \geq m_R\vartheta^k \tag{3.26}$$

with $m_R \in (0.5,1)$. Set

$$u^{k+1}_{int} = 2u^k(1 - [f_l(\widehat{X}^k;\hat{y}^k) - f_u(X^{k+1};y^{k+1})]/\vartheta^k), \tag{3.27}$$

then $u^{k+1}_{int}/u^k \leq 2(1-m_R) < 1$ if (3.26) satisfies. In this setting, we set

$$u^{k+1} = \max\left\{u^{k+1}_{int}, u^k/10, u_{min}\right\}, \tag{3.28}$$

where u_{min} is a small positive constant.

Now, we investigate the case when u_k seems to be small.

By (3.15), we obtain

$$\vartheta^k = \tau^k - a\hat{\epsilon}^k + u_k l^k_d. \tag{3.29}$$

We define the error of new approximation by

$$e^k_v = f_u(\widehat{X}^k;\hat{y}^k) - f_{W^{k+1}_{ap}}(\widehat{X}^k;\hat{y}^k) \tag{3.30}$$

and the variation

$$(var)_k = f(\widehat{X}^k;\hat{y}^k) - \min\left\{f(X;y)\middle|(X;y) \in \mathscr{B}_{\mathcal{M}}(\widehat{X}^k;\hat{y}^k,l^k_d)\right\},$$

where $\mathscr{B}_{\mathcal{M}}(\widehat{X}^k;\hat{y}^k,l^k_d) = \left\{(X;y) \in \mathcal{S}^n \times \mathbb{R}^m \middle| \|X-\widehat{X}^k\|^2_{\mathcal{M}_{X,k}} + \|y-\hat{y}^k\|^2_{\mathbf{M}_{y,k}} \leq l^k_d\right\}$.

Thus, by virtue of $W^{k+1} \in \widehat{\mathcal{W}}^k$, we obtain

$$\begin{aligned} 0 \leq &(var)_k \leq f(\widehat{X}^k;\hat{y}^k) - \min\left\{f_{\widehat{\mathcal{W}}^k}(X;y)\middle|(X;y) \in \mathscr{B}_{\mathcal{M}}(\widehat{X}^k;\hat{y}^k,l^k_d)\right\} \\ &\leq f(\widehat{X}^k;\hat{y}^k) - \min\left\{f_{W^{k+1}}(X;y)\middle|(X;y) \in \mathscr{B}_{\mathcal{M}}(\widehat{X}^k;\hat{y}^k,l^k_d)\right\} \\ &\leq f_l(\widehat{X}^k;\hat{y}^k) - f_{W^{k+1}}(X^{k+1};y^{k+1}) + a\hat{\epsilon}^k = \vartheta^k + a\hat{\epsilon}^k. \end{aligned}$$

One may use the test

$$e^k_v > \vartheta^k + a\hat{\epsilon}^k \tag{3.31}$$

to decide that u_k should be increased.

3.4. **The inexact spectral bundle algorithm.** In this subsection, we present the inexact spectral bundle method for the eigenvalue optimization problem (Eigform), combining ideas from the algorithms in [24], [34] and [14].

In what follows, the expression $[\lambda_{ap}, v_{ap}, \epsilon] = \textbf{eigest}(X, y, \delta)$ is the routine that produces a *Ritz triplet* $(\lambda_{ap}, v_{ap}, \epsilon)$ for $\lambda_1(-C - \mathcal{A}^*(y) - \mathcal{Q}(X))$ such that $a\epsilon \le \delta$.

Algorithm 3.1. Input: *An initial point* $(X^0; y^0) \in \mathcal{S}^n \times \mathbb{R}^m$, *an* $\varepsilon \ge 0$ *for termination, an improvement parameter* $m_L \in (0, 0.5)$, *an initial subdifferential tolerance* δ^0, *an initial weight* $u_0 > \|\mathcal{Q}\|$, *a minimal weight* $u_{min} > \|\mathcal{Q}\|$, *set the variation estimate* $\theta_v^0 = \infty$.

*1.***(Initialization)** $k = 0$, $\widehat{X}^0 = X^0$, $\hat{y}^0 = y^0$, $s = 0$. *Call* $[\hat{\lambda}^0_{ap}, \hat{v}^0_{ap}, \hat{\epsilon}^0] = \textbf{eigest}(\widehat{X}^0, \hat{y}^0, \delta^0)$. *Compute* $(g^0_X; g^0_y)$, $\widehat{\mathcal{W}}^0$ *and* $f_l(\widehat{X}^0; \hat{y}^0)$.

*2.***(Trial point finding)** *Solve (SQSDP) to get* W^{k+1}, *compute* $(X^{k+1}; y^{k+1})$ *and* l^k_d *from (3.5) and (3.16), respectively.*

*3.***(Evaluation)** *Call* $[\lambda^{k+1}_{ap}, v^{k+1}_{ap}, \epsilon^{k+1}] = \textbf{eigest}(X^{k+1}, y^{k+1}, \delta^k)$. *Compute* η^k, τ^k *and* ν^k *from (3.14), (3.15) and (3.19) respectively. Set* $W^{k+1}_{ap} = a v^{k+1}_{ap}(v^{k+1}_{ap})^T$.

*4.***(Stopping criterion)** *If* $\nu^k \le \varepsilon$, *then stop.*

*5.***(Descent step)** *If* $f_l(\widehat{X}^k; \hat{y}^k) - f_u(X^{k+1}; y^{k+1}) \ge m_L \vartheta^k$, *then set* $(\widehat{X}^{k+1}; \hat{y}^{k+1}) = (X^{k+1}; y^{k+1})$, $\hat{\lambda}^{k+1}_{ap} = \lambda^{k+1}_{ap}$, $\hat{\epsilon}^{k+1} = \epsilon^{k+1}$, *and* $\hat{v}^{k+1}_{ap} = v^{k+1}_{ap}$, *select* $u_{k+1} \in [u_{min}, u_k]$*(e.g. by (3.28)) and set*

$$\theta_v^{k+1} = \max\left\{\theta_v^k, 0.5(\vartheta^k + a\hat{\epsilon}^k)\right\}, \tag{3.32}$$

set $\delta^{k+1} \le \delta^k$ *to ensure convergence, set* $s = s + 1$, *choose the operator* $\mathcal{M}_{X,k+1}$ *and the matrix* $\mathbf{M}_{y,k+1}$ *such that (1.3) holds and continue with step 7. Otherwise continue with step 6.*

*6.***(Null step)** *Set* $(\widehat{X}^{k+1}; \hat{y}^{k+1}) = (\widehat{X}^k; \hat{y}^k)$. *If* $a\hat{\epsilon}^k > 10\delta^k$ *then call* $[\hat{\lambda}^{k+1}_{ap}, \hat{v}^{k+1}_{ap}, \hat{\epsilon}^{k+1}] = \textbf{eigest}(\widehat{X}^k, \hat{y}^k, \delta^k)$, *else set* $\hat{\lambda}^{k+1}_{ap} = \hat{\lambda}^k_{ap}$, $\hat{\epsilon}^{k+1} = \hat{\epsilon}^k$, *and* $\hat{v}^{k+1}_{ap} = \hat{v}^k_{ap}$. *Set*

$$\theta_v^{k+1} = \min\left\{\theta_v^k, \vartheta^k + a\hat{\epsilon}^k\right\} \tag{3.33}$$

and either set $u_{k+1} = u_k$ *or choose* $u_{k+1} \in [u_k, 10u_k]$*(e.g.* $u^{k+1} = \min\left\{u^{k+1}_{int}, 10u^k\right\}$*) if*

$$e_v^k > \theta_v^{k+1} \tag{3.34}$$

Set $\delta^{k+1} \le \delta^k$ *to ensure convergence. Choose the operator* $\mathcal{M}_{X,k+1}$ *and the matrix* $\mathbf{M}_{y,k+1}$ *such that (1.4) satisfies.*

*7.***(Updating** $\widehat{\mathcal{W}}^k$**)** *Choose a* $\widehat{\mathcal{W}}^{k+1} \supset \left\{W^{k+1}, W^{k+1}_{ap}\right\}$ *of the form (3.1).*

8. k:=k+1 and goto step 2.

Remark 3.2. (1). Set $P_0 = \hat{v}^0_{ap}$, $\overline{W}_0 = a\hat{v}^0_{ap}(\hat{v}^0_{ap})^T$, then compute $\widehat{\mathcal{W}}^0$ by the form (3.1).

(2). We can choose $\widehat{\mathcal{W}}^{k+1}$ as follows. Decompose V^* into $V^* = R_1\Lambda_1 R_1^T + R_2\Lambda_2 R_2^T$ with $rank(R_1) \le r - 1$. Compute $\overline{W}_{k+1}$ and P_{k+1} by using (3.23) and (3.22), respectively, where an upper bound $r \ge 1$ on the number of columns of P.

(3). It has been pointed out in [5, Sect.10] that a sequence of null steps between two consecutive stability centers is just for the sake of the improvement of the model but possessing no new reliable information, the operator $\mathcal{M}_{X,k}$ and the matrix $\mathbf{M}_{y,k}$ are only updated when a descent step comes up. This updating rule of the metric originates from [27], and it is employed in [23]. At the same time, to the best of our knowledge, most of the literature about bundle methods exploit the special matrix $\mathbf{M}_{y,k} = I$ and exclude the matrix variable X, for instance, see [2, 4, 13, 14, 16, 24, 34, 44]. However, in order to ensure convergence, we should use variable metric at both descent steps and null steps in Algorithm 3.1.

(4). Henceforth, we presume that $\{\|\mathcal{M}_{X,k}\|, \|M_{y,k}\|\}$ is bounded and any accumulation point of $\{\mathcal{M}_{X,k}\}$ is still a self-adjoint positive definite linear operator, and any accumulation point of $\{\mathbf{M}_{y,k}\}$ remains positive. As a consequence, $\{c_k\}$ is bounded. More precisely, we take the following assumption:

(A3) If Algorithm 3.1 executes an infinite number of null steps or an infinite number of descent steps, then $\{\mathcal{M}_{X,k}\}$ converges a self-adjoint positive definite linear operator and $\{\mathbf{M}_{y,k}\}$ converges to a positive definite matrix. □

3.5. **Convergence analysis.** In this subsection, we will prove the convergence of Algorithm 3.1 for $\varepsilon = 0$. If Algorithm 3.1 terminates after a finite number K of iterations, then by Proposition 3.3, $(\widehat{X}^K; \hat{y}^K)$ is optimal. Thus the case of interest is that Algorithm 3.1 does not stop. In showing that the convergence of the algorithm in this setting, we will employ several propositions. Note that the essential ideas of their proofs are very similar to those presented in [14], so their proofs are given in Appendix A, but the approach will be used in this section. First we consider the case when only null steps occur after some iteration N.

Proposition 3.5. *Assume that, starting with iteration N, the implementation of Algorithm 3.1 is without the descent test, i.e. $\widehat{X} = \widehat{X}^N = \widehat{X}^{N+1} = \cdots$, $\hat{y} = \hat{y}^N = \hat{y}^{N+1} = \cdots$. For any $k > N$, set*

$$X^k = X^N_{min}(W^k) = \widehat{X} + \mathcal{T}^{-1}_{k-1}(\mathcal{Q}(W^k - \widehat{X})), \tag{3.35}$$

$$y^k = y^N_{min}(W^k) = \hat{y} + \frac{1}{u_{k-1}}\mathbf{M}^{-1}_{y,k-1}(\mathcal{A}(W^k) - b). \tag{3.36}$$

*Suppose that **(A3)** holds and $\delta^k \to 0$, then $(X^k; y^k) \to (\widehat{X}; \hat{y}) \in arg\min_{X,y} f(X; y)$ and $\nu^k \to 0$ as $k \to \infty$.*

The preceding result is with respect to convergence of Algorithm 3.1 when the number s of descent steps is bounded. It remains to analyze the case of a unbounded s. Over the course of implementing the algorithm, it may happen that null steps appear between two consecutive descent steps. At this point, we would like to investigate an infinite number of descent steps, for the sake of convenience, we can discard all null steps. In other words, we may focus on the situation $(\widehat{X}^k; \hat{y}^k) = (X^k; y^k)$. Thus we assign the remaining iterates and the corresponding W^k with a new index h, and assume that, for any h,

$$X^{h+1} = X^h_{min}(W^{h+1}) = X^h + \mathcal{T}^{-1}_h(\mathcal{Q}(W^{h+1} - X^h)), \tag{3.37}$$

$$y^{h+1} = y^h_{min}(W^{h+1}) = y^h + \frac{1}{u_h}\mathbf{M}^{-1}_{y,h}(\mathcal{A}(W^{h+1}) - b), \tag{3.38}$$

$$f_l(X^h; y^h) - f_u(X^{h+1}; y^{h+1}) \geq m_L(f_l(X^h; y^h) - f_{W^{h+1}}(X^{h+1}; y^{h+1})). \tag{3.39}$$

Moreover, if no null steps occur, then

$$\begin{aligned} f_l(X^h; y^h) - m_L(f_l(X^h; y^h) - f_{W^{h+1}}(X^{h+1}; y^{h+1})) \geq f_u(X^{h+1}; y^{h+1}) \geq f(X^{h+1}; y^{h+1}) \\ \geq f_{W^{h+1}}(X^{h+1}; y^{h+1}), \end{aligned}$$

whence $f_l(X^h; y^h) - f_u(X^{h+1}; y^{h+1}) \geq m_L(f_l(X^h; y^h) - f_{W^{h+1}}(X^{h+1}; y^{h+1})) \geq 0$. Therefore,

$$f(X^h; y^h) - f(X^{h+1}; y^{h+1}) \geq f_l(X^h; y^h) - f_u(X^{h+1}; y^{h+1}) \geq 0. \tag{3.40}$$

Owing to (3.40), the sequence $\{f(X^h; y^h)\}_{h=1}^{\infty}$ is nonincreasing, hence $f(X^h; y^h) \downarrow \inf f$ as $h \to \infty$. For the case of interest, we make the following assumption

$$\exists\ (\widetilde{X}; \tilde{y}) \in \mathcal{S}^n \times \mathbb{R}^m,\ s.t.\ f(X^h; y^h) \geq f(\widetilde{X}; \tilde{y})\ \forall\ h. \tag{3.41}$$

Proposition 3.6. *Assume that both (3.41) and* **(A3)** *satisfy, then* $(X^h; y^h)$ *converges to a minimizer of* $f(X; y)$ *and* $\lim_{h\to\infty} \nu^h = 0$.

We summarize the above discussion in the following theorem.

Theorem 3.1. *Let* $\{(\widehat{X}^k; \hat{y}^k)\}$ *be the sequence of points generated by Algorithm 3.1. Suppose that* **(A3)** *satisfies and* $\lim_{k\to\infty} \delta^k = 0$*, then either* $(\widehat{X}^k; \hat{y}^k) \to (\bar{X}; \bar{y}) \in arg\min_{X,y} f(X; y)$ *or there is no minimizer and* $\|(\widehat{X}^k; \hat{y}^k)\| \to \infty$ *as* $k \to \infty$*. In both cases,* $f(\widehat{X}^k; \hat{y}^k) \downarrow \inf f(X; y)$ *as* $k \to \infty$.

Proof. The proof is straightforward by Proposition 3.5 and Proposition 3.6. □

3.6. **Approximate solutions and error bound.** In most cases, it may be tremendously expensive to pursue an exact optimal solution. It should also be realized that many mathematical programs do not have an exact optimal solution. Meanwhile, most numerical methods attempting to solve global optimization only yield approximate optimal solutions. Therefore, it is sensible meaning to resort approximate solutions.

At this point, we need to recall definitions of approximate solutions.

Definition 3.1. [16] Let $\epsilon, \epsilon_s > 0$. The point $(X^*; y^*) \in \mathcal{S}^n \times \mathbb{R}^m$ is called an ϵ**-**ϵ_s**-optimal** solution of (EigForm) if

$$f(X^*; y^*) \le f(X; y) + \epsilon(\|X - X^*\|^2 + \|y - y^*\|^2)^{1/2} + (\epsilon + \epsilon_s), \forall\ (X; y) \in \mathcal{S}^n \times \mathbb{R}^m.$$

Hence, $\epsilon = 0$ implies ϵ_s**-optimality** of $(X^*; y^*)$ in the sense of [30], while $\epsilon_s = 0$ yields ϵ**-optimality** of $(X^*; y^*)$ in the sense of [44]. Henceforth we call them ϵ_s^{F1}**-optimal** solution and ϵ^{F2}**-optimal** solution, respectively. Plainly, by virtue of (3.17) and (3.18), if $\vartheta^k \le \epsilon$, then either $(\widehat{X}^k; \hat{y}^k)$ or $(X^{k+1}; y^{k+1})$ is an ϵ^{F2}**-optimal** solution. We would like to find an ϵ^{F1}**-optimal** solution in the following.

From the convergence analysis of Algorithm 3.1, we see that the rule of choosing δ^k should ensure that $\delta^k \downarrow 0$ in the case of a bounded s, which counts descent steps. However, we could wish that $\delta^k \downarrow \delta > 0$, which is easy to be controlled in practice. To this end, we take the following assumptions, which are tractable in practice.

(A4) $a\lambda_{ap}^k \ge \langle W^k, -C - \mathcal{A}^*(y^k) - \mathcal{Q}(X^k)\rangle$ for all k;

(A5) In the context of Proposition 3.5 and $\delta^k \downarrow \delta > 0$,

$$a\hat{\lambda}_{ap}^k - \frac{2\delta}{1 - m_L} \ge \langle W^k, -C - \mathcal{A}^*(\hat{y}) - \mathcal{Q}(\widehat{X})\rangle,\ \forall k > N. \tag{3.42}$$

Note that, (3.42) yields

$$f(\widehat{X}; \hat{y}) - \frac{2\delta}{1 - m_L} \ge f_{W^k}(\widehat{X}; \hat{y}),\ \forall k > N. \tag{3.43}$$

By Proposition 3.3, for $k > N$,

$$\begin{aligned} &f(\widehat{X}; \hat{y}) \ge f_{W^k}(\widehat{X}; \hat{y}) = L^{k-1}(\widehat{X}, \hat{y}, W^k) \\ &= L^{k-1}(X^k, y^k, W^k) + \frac{u_{k-1}}{2}(\|\widehat{X} - X^k\|^2_{\mathcal{M}_{X,k-1}} + \|\hat{y} - y^k\|^2_{\mathbf{M}_{y,k-1}}) \ge f_{W^k}(X^k; y^k), \end{aligned}$$

hence, if $f(\widehat{X}; \hat{y}) = f_{W^k}(X^k; y^k)$, then $(\widehat{X}; \hat{y})$ is optimal. Hence the case of interest is that $f(\widehat{X}; \hat{y}) > f_{W^k}(X^k; y^k)$, which may follow from (3.43). Thus **(A5)** is possible.

The next theorem gives the counterpart in Proposition 3.5.

Proposition 3.7. *In the context of Proposition 3.5, suppose that both* **(A4)** *and* **(A5)** *hold, and at step 5 of Algorithm 3.1 we choose $u_{k+1} \in [u_k, u_{max}]$ at null steps, where $u_{max} > 0$ is fixed. Then $(X^k; y^k) \to (\widehat{X}; \hat{y})$ as $k \to \infty$, which is $(\frac{2\delta}{1-m_L})^{F1}$*-**optimal** *solution of (Eigform), and $\overline{\lim}_{k\to\infty}\nu^k \leq 2\delta$.*

Proof. By the proof of Proposition 3.5, we see that

$$\overline{\lim}_{k\to\infty}(f_l(X^k; y^k) - f_{W^k}(X^k; y^k)) \leq \overline{\lim}_{k\to\infty}(f(X^k; y^k) - f_{W^k}(X^k; y^k)) \leq \delta. \tag{3.44}$$

On the one hand, if no descent steps occur, then

$$\begin{aligned} f_l(\widehat{X}^{k-1}; \hat{y}^{k-1}) - f_u(X^k; y^k) &< m_L(f_l(\widehat{X}^{k-1}; \hat{y}^{k-1}) - f_{W^k}(X^k; y^k)) \\ &\leq m_L(f_l(\widehat{X}^{k-1}; \hat{y}^{k-1}) + a\hat{\epsilon}^{k-1} - f_{W^k}(X^k; y^k)). \end{aligned}$$

The above inequalities can be rewritten as

$$f_l(X^k; y^k) + a\epsilon^k + a\hat{\epsilon}^{k-1} - f_{W^k}(X^k; y^k) \geq (1 - m_L)(f_l(\widehat{X}^{k-1}; \hat{y}^{k-1}) + a\hat{\epsilon}^{k-1} - f_{W^k}(X^k; y^k)).$$

Moreover, (3.43) and the inequalities followed yield $f(\widehat{X}; \hat{y}) - \frac{2\delta}{1-m_L} \geq f_{W^k}(X^k; y^k)$, $\forall k > N$.

For $k > N$, **(A4)** implies that $f_l(X^k; y^k) \geq f_{W^k}(X^k; y^k)$.

On the other hand, by step 6,

$$a\hat{\epsilon}^k \leq \min\{\delta^{k-1}, 10\delta^k\}. \tag{3.45}$$

Hence, for $k > N$, $f_l(X^k; y^k) + \delta^{k-1} + \delta^{k-2} - f_{W^k}(X^k; y^k) \geq f_l(X^k; y^k) + a\epsilon^k + a\hat{\epsilon}^{k-1} - f_{W^k}(X^k; y^k) \geq 2\delta$, then $(1 - m_L)(f(\widehat{X}; \hat{y}) - f_{W^k}(X^k; y^k)) \to 2\delta$ as $k \to \infty$.

Using (3.43), we obtain $f(\widehat{X}; \hat{y}) = \max_{W\in a\mathcal{W}} f_W(\widehat{X}; \hat{y}) \geq f_{W^k}(\widehat{X}; \hat{y}) + \frac{2\delta}{1-m_L}$. Applying the same argument used in the end of the proof of Proposition 3.5, we obtain

$$\begin{aligned} &\lim_{k\to\infty}(g_X^{k-1}; g_y^{k-1}) = (0; 0) \\ &(X^k; y^k) \to (\widehat{X}; \hat{y}) \text{ as } k \to \infty; \\ &f(X; y) \geq f(\widehat{X}; \hat{y}) - \frac{2\delta}{1 - m_L} \text{ for any } (X; y) \in \mathcal{S}^n \times \mathbb{R}^m. \end{aligned} \tag{3.46}$$

Since $\eta^k = f_l(X^k; y^k) - f_{W^k}(X^k; y^k) + a\epsilon^k$, together with (3.44) and (3.45), then

$$\overline{\lim}_{k\to\infty}\eta^k \leq 2\delta. \tag{3.47}$$

Therefore, putting together (3.47) and (3.46), we obtain $\overline{\lim}_{k\to\infty}\nu^k \leq 2\delta$. □

In short, the next result sharpens the convergence behavior of the Algorithm 3.1 under the inexact setting.

Theorem 3.2. *Suppose that* **(A3)**, **(A4)** *and* **(A5)** *hold, and at step 5 of Algorithm 3.1 we choose $u_{k+1} \in [u_k, u_{max}]$ at null steps, where $u_{max} > 0$ is fixed. Let $\{(\widehat{X}^k; \hat{y}^k)\}$ be the sequence of points generated by Algorithm 3.1. There are three possible cases :*

1.If the algorithm terminates after a finite number K of iterations, $(\widehat{X}^K; \hat{y}^K)$ is optimal.

2. If the algorithm does not terminates and $s \to \infty$, then either $(\widehat{X}^k; \hat{y}^k) \to (\bar{X}; \bar{y}) \in \arg\min_{X,y} f(X; y)$ or there is no minimizer and $\|(\widehat{X}^k; \hat{y}^k)\| \to \infty$ as $k \to \infty$.

3. If the algorithm does not terminates but s is bounded, then $(\widehat{X}^k; \hat{y}^k) \to (\widehat{X}; \hat{y})$, which is $(\frac{2\delta}{1-m_L})^{F1}$-**optimal** *solution of (Eigform).*

Proof. The proof is straightforward by Theorem 3.1 and Proposition 3.7. □

So far, we have investigated the algorithm and its convergence analysis for (EigForm). It is natural to ask whether the same algorithm can be applied to handle (P). Thanks to Proposition 2.1, we can relate the accumulation points of $\{W^k\}$ to the optimal or approximate solutions of (P).

Theorem 3.3. *Let $\mathcal{O}_P = arg\min\{\langle X, \mathcal{Q}(X)\rangle/2 + \langle C, X\rangle \big| \mathcal{A}(X) = b, X \succeq 0\}$. Assume that $arg\min_{X,y} f(X; y) \neq \emptyset$,* **(A1)**, **(A2)** *and* **(A3)** *satisfy, then we have*

(i) if k is finite with $k = K$ on termination, then $W^{K+1} \in \mathcal{O}_P$,

*(ii) if there is an infinite number of iterations and no descent steps occur after iteration N, then each accumulation point of $\{W^k\}$ is a $(\frac{2\delta}{1-m_L})^{F1}$-**optimal** solution of (P) in the context of Proposition 3.7.*

(iii) if there is an infinite number of descent steps, then all accumulation points of $\{W^h\}$ giving rise to descent steps lie in $\mathcal{O}_P$.

Proof. Set $p := \min\left\{h(X)\big|\mathcal{A}(X) = b, X \succeq 0\right\}, -d := \min\left\{g(X; y)\big|\mathcal{Q}(X) + \mathcal{A}^*(y) + C \succeq 0\right\}$, then the assumption **(A1)**, $arg\min_{X,y} f(X; y) \neq \emptyset$ and Proposition 2.1 imply $p = d$.

Let $\{(\widehat{X}^k; \hat{y}^k)\}$ be a sequence of points generated by Algorithm 3.1.

(i) If k is finite with $k = K$ on termination, by Proposition 3.3, it holds $(X^{K+1}; y^{K+1}) = (\widehat{X}^K; \hat{y}^K)$ and $(X^{K+1}; y^{K+1})$ is optimal, thus

$$\mathcal{Q}(X^{K+1} - W^{K+1}) = 0 \text{ and } b - \mathcal{A}(W^{K+1}) = 0. \tag{3.48}$$

Note that $W^{K+1} \in a\mathcal{W}$, hence it is feasible for (P).

On the other hand, due to (3.48) , Proposition 3.3 and Proposition 2.1, we have

$$\begin{aligned}
-h(W^{K+1}) &= -\frac{1}{2}\langle W^{K+1}, \mathcal{Q}(W^{K+1})\rangle - \langle C, W^{K+1}\rangle = -\frac{1}{2}\langle W^{K+1}, \mathcal{Q}(X^{K+1})\rangle - \langle C, W^{K+1}\rangle \\
&= \langle -C - \mathcal{A}^*(y^{K+1}) - \mathcal{Q}(X^{K+1}), W^{K+1}\rangle + b^T y^{K+1} + \frac{1}{2}\langle X^{K+1}, \mathcal{Q}(X^{K+1})\rangle \\
&= f(\widehat{X}^K; \hat{y}^K) = g(\widehat{X}^K; \hat{y}^K) = -d,
\end{aligned}$$

whence $W^{K+1} \in \mathcal{O}_P$.

(ii) Since $W^{k+1} \in \widehat{\mathcal{W}}^k \subseteq a\mathcal{W}$ and $a\mathcal{W}$ is compact, $\{W^k\}$ has accumulation points and they are contained in $a\mathcal{W}$.

Assume that a subsequence $\{W^{k_j}\}$ converges $W_* \in a\mathcal{W}$. By the proof of Proposition 3.7, $X^{k_j} \to \widehat{X}^N, y^{k_j} \to \hat{y}^N$ and $f_{W^{k_j}}(X^{k_j}; y^{k_j}) \to f(\widehat{X}^N; \hat{y}^N) - \frac{2\delta}{1-m_L}$, $b - \mathcal{A}(W^{k_j}) \to 0$, $\mathcal{Q}(\widehat{X}^N - W^{k_j}) \to 0$, whence

$$\mathcal{Q}(\widehat{X}^N - W_*) = 0,\ b - \mathcal{A}(W_*) = 0. \tag{3.49}$$

Therefore, we have

$$\begin{aligned}
f_{W^{k_j}}(X^{k_j}; y^{k_j}) &= \langle -C, W^{k_j}\rangle + \langle b - \mathcal{A}(W^{k_j}), y^{k_j}\rangle + \langle X^{k_j}, \mathcal{Q}(\widehat{X}^N - W^{k_j})\rangle - \frac{1}{2}\langle \widehat{X}^N, \mathcal{Q}(\widehat{X}^N)\rangle \\
&+ \frac{1}{2}\langle X^{k_j} - \widehat{X}^N, \mathcal{Q}(X^{k_j} - \widehat{X}^N)\rangle \to \langle -C, W_*\rangle - \frac{1}{2}\langle \widehat{X}^N, \mathcal{Q}(\widehat{X}^N)\rangle \\
&= \langle -C, W_*\rangle - \frac{1}{2}\langle W_*, \mathcal{Q}(W_*)\rangle = -h(W_*) = f(\widehat{X}^N; \hat{y}^N) - \frac{2\delta}{1 - m_L},
\end{aligned}$$

then $-h(W_*) + \frac{2\delta}{1-m_L} = f(\widehat{X}^N; \hat{y}^N) \geq -d = -p \geq -h(X)$ for any feasible solution X for (P), the first inequality results from Proposition 2.1. Together with (3.49), the desired result holds.

(iii) Since $arg\min_{X,y} f(X; y) \neq \emptyset$, (3.41) satisfies. By Proposition 3.6, following the same approach we have done in the proof of (ii), we get the desired result. ∎

In practice, the algorithm will cease at some iteration k due to the given stopping criteria. Hence we could wish the current stability center $(\widehat{X}^k; \hat{y}^k)$ to be an δ^{F1}**-optimal** solution. To this end, we now highlight its sufficient conditions, in which cases the current stability center $(\widehat{X}^k; \hat{y}^k)$ is already δ^{F1}**-optimal**. The following result is analogous to that in [25, LEMMA 2.3].

Proposition 3.8. *i) If* $f_l(\widehat{X}^k; \hat{y}^k) \le \inf f_{\widehat{\mathcal{W}}^k}$ *and* $\delta^k \downarrow \delta > 0$, *then* $f(\widehat{X}^k; \hat{y}^k) \le \inf f + \delta$.
ii) Suppose that **(A4)** *holds, then* $\nu^k = 0$ *implies* $f_l(\widehat{X}^k; \hat{y}^k) \le \min f_{\widehat{\mathcal{W}}^k} = f_{\widehat{\mathcal{W}}^k}(\widehat{X}^k; \hat{y}^k)$.

Proof. i) In view of $f_{\widehat{\mathcal{W}}^k}(X; y) \le f(X; y)$ for all $(X; y) \in \mathcal{S}^n \times \mathbb{R}^m$, it holds $\inf f_{\widehat{\mathcal{W}}^k} \le \inf f$.
On the other hand, $f(\widehat{X}^k; \hat{y}^k) - a\hat{\epsilon}^k \le f_l(\widehat{X}^k; \hat{y}^k)$. Therefore $f_l(\widehat{X}^k; \hat{y}^k) \le \inf f_{\widehat{\mathcal{W}}^k}$ yields $f(\widehat{X}^k; \hat{y}^k) \le \inf f + a\hat{\epsilon}^k \le \inf f + \delta$.
ii) By **(A4)**, $\eta^k > 0$. Thus $\nu^k = 0$ results in $(X^{k+1}; y^{k+1}) = (\widehat{X}^k; \hat{y}^k)$ and $\tau^k \le 0$. Consequently, $(g_X^k; y_y^k) = (0; 0)$ and $f_l(\widehat{X}^k; \hat{y}^k) = \bar{f}_{\mathcal{W}}^k(\widehat{X}^k; \hat{y}^k) + (\tau^k - a\hat{\epsilon}^k) \le \bar{f}_{\mathcal{W}}^k(\widehat{X}^k; \hat{y}^k)$.
Due to (3.9), $\min f_{\widehat{\mathcal{W}}^k} = f_{\widehat{\mathcal{W}}^k}(X^{k+1}; y^{k+1}) = f_{\widehat{\mathcal{W}}^k}(\widehat{X}^k; \hat{y}^k)$ since $(g_X^k; y_y^k) = (0; 0)$.
On the other hand, (3.10) implies $f_{\widehat{\mathcal{W}}^k}(X^{k+1}; y^{k+1}) = \bar{f}_{\mathcal{W}}^k(X^{k+1}; y^{k+1})$. Therefore, the desired result holds. □

Remark 3.3. **(A4)** is always true in our numerical experiments in Section 4. □

We close this section with a global error bound for (EigForm), followed by an interpretation that the optimality measure (3.19) makes sense in practice.

Henceforth, we assume that $f_* = \inf_{X;y} f(X; y) > -\infty$.

At this point, we need to recall two existing results which play a key role in our conclusions. We only quote the results in our context.

Theorem 3.4. *[52, Theorem 2.2] Suppose that, for some* $(X_0; y_0) \in \mathcal{S}^n \times \mathbb{R}^m$, $0 < \zeta \le +\infty$, *and* $0 < \mu < +\infty$, *and* $0 < \gamma \le \zeta/(2\mu)$, *the set* $T = B((X_0; y_0), \gamma/2) \cap \{(X; y) \in \mathcal{S}^n \times \mathbb{R}^m | f(X; y) - f_* < \gamma\}$ *is nonempty and for all* $(g_X; g_y) \in \partial f(X; y)$ *and each* $(X; y) \in T_1$, *where* $T_1 = B((X_0; y_0), \gamma/2) \cap \{(X; y) \in \mathcal{S}^n \times \mathbb{R}^m | 0 < f(X; y) - f_* < \gamma\}$, *it holds* $\|(g_X; g_y)\| \ge \mu^{-1}$, *then* $\mathcal{O}_* := \arg\min_{X,y} f(X; y)$ *is nonempty and*

$$d((X; y), \mathcal{O}_*) \le \mu(f(X; y) - f_*) \text{ for all } (X; y) \in T.$$

Moreover, if $(X_0; y_0) \in \mathcal{O}_*$, *then the condition* $0 < \gamma \le \zeta/(2\mu)$ *can be replaced with* $0 < \gamma \le +\infty$.

Proposition 3.9. *[22, Lemma 3] Let* $p : \mathbb{R}^n \to \mathbb{R}$ *be a locally Lipschitz mapping. Suppose that:*
(i) there exist $\sigma > 0$ *and* $W \in \mathcal{S}^n$ *such that*

$$\sigma I - W \in \mathcal{S}_-^n, \tag{3.50}$$

(ii) $\lambda_1(\sigma I - W) \sum_{\mu_i \ge 0} \mu_i + \lambda_n(\sigma I - W) \sum_{\mu_i \le 0} \mu_i \le 0$, *for any* $\mu \in \partial p(\lambda(X))$.
Hence for each $A \in \partial(p \circ \lambda)(X)$, *there is* $\mu \in \partial p(\lambda(X))$ *such that* $\langle A, W\rangle \ge \sigma \sum_{1=1}^n \mu_i$.

It is known that (3.50) is a Slater type condition. When applied to $J\Phi(X; y)(D; d)$, it amounts to that there is some $(D; d) \in \mathcal{S}^n \times \mathbb{R}^m$ such that $J\Phi(X; y)(D; d)$ is positive definite for all $(X; y) \in \mathcal{S}^n \times \mathbb{R}^m$. A global error bound for the special case when $\Phi(X; y) = -C - \mathcal{A}^*(y)$ under this type condition is given in [22, Corollary 2]. Using the similar arguments, we can derive the following global error bound for (Eigform).

Theorem 3.5. *Suppose that* **(A1)** *and* **(A2)** *satisfy. For any finite $N>0$, let $\{(X^k;y^k)\}_{k=1}^N$ be a finite sequence of points generated by Algorithm 3.1. Further, suppose that there exist $\sigma>0$ and $(D;d)\in\mathcal{S}^n\times\mathbb{R}^m$ such that*

$$\sigma I - J\Phi(X^k;y^k)(D;d)\in\mathcal{S}^n_- \text{ for any } k=1,\cdots,N. \tag{3.51}$$

Thus $\mathcal{O}_*$ is nonempty and

$$d((X^k;y^k),\mathcal{O}_*)\le\frac{\sqrt{\|D\|^2+\|d\|^2}}{\sigma}(f(X^k;y^k)-f_*) \text{ for any } k=1,\cdots,N. \tag{3.52}$$

Proof. After going through the proof of Proposition 2.1, we see that

$$\mathcal{O}_D := arg\min\left\{g(X;y)\ \middle|\ \mathcal{Q}(X)+\mathcal{A}^*(y)+C\succeq 0\right\},$$
$$\mathcal{O}_* = \left\{(X^*;y^*)+t(0;\bar{y})\ \middle|\ (X^*;y^*)\in\mathcal{O}_D, t\in\mathbb{R}\right\}.$$

Plainly, **(A1)** and $f_*>-\infty$ imply that $\mathcal{O}_D$ is nonempty, whence $\mathcal{O}_*$ is also nonempty with the aid of **(A2)**. Then one can choose the desired point in Theorem 3.4 such that $(X_0;y_0)\in\mathcal{O}_*$, hence it may be $\gamma=\infty$ by Theorem 3.4 so that $\{(X^k;y^k)\}_{k=1}^N\subseteq T$.

By setting $p(x_1,\cdots,x_n)=x_1-f_*$, we have $f(X,y)-f_*=(p\circ\lambda\circ\Phi)(X;y)$. It is easy to see that $\partial p(\lambda(\Phi(X;y)))=\{e_1\}$ for all $(X;y)\in\mathcal{S}^n\times\mathbb{R}^m$. On the other hand, for any $(g_{X^k};g_{y^k})\in\partial f(X^k;y^k), k=1,\cdots,N$, via Theorem 2.1, there is $G^k\in\partial(p\circ\lambda)(\Phi(X^k;y^k))$ such that $(g_{X^k};g_{y^k})=J\Phi(X^k;y^k)^*(G^k)$. Using (3.51), by Proposition 3.9, we have $\langle(g_{X^k};g_{y^k}),(D;d)\rangle=\langle J\Phi(X^k;y^k)^*(G^k),(D;d)\rangle=\langle G^k,J\Phi(X^k;y^k)(D;d)\rangle\ge\sigma$, hence $\|(g_{X^k};g_{y^k})\|\ge\frac{\sigma}{\|(D;d)\|}$ and the proof is completed. □

We now return to the construction of the optimality measure ν^k. Suppose that the conditions in Theorem 3.5 hold and at iteration $k+1$ it satisfies $\nu^{k+1}\le\varepsilon$. Putting together (3.17) and (3.18), via Theorem 3.5 and the fact that $\{(\widehat{X}^k;\hat{y}^k)\}$ is a subsequence of $\{(X^k;y^k)\}$, we obtain that one of the following inequalities is satisfied

$$d((\widehat{X}^k;\hat{y}^k),\mathcal{O}_*)\le\frac{\varepsilon}{\frac{\sigma}{\sqrt{\|D\|^2+\|d\|^2}}-\varepsilon} \tag{3.53}$$

$$d((X^{k+1};y^{k+1}),\mathcal{O}_*)\le\frac{\varepsilon}{\frac{\sigma}{\sqrt{\|D\|^2+\|d\|^2}}-\varepsilon}. \tag{3.54}$$

The following example says that finding the proposed σ and $(D;d)$ in the above theorem is a possible task.

Example 3.1. *We continue Example 2.1. We show that Problem (2.1) satisfies condition (3.51) for the special case where $\mathcal{Q}$ has the property that there is some scalar ρ with $\rho<\|\mathcal{Q}\|$ such that $tr(\mathcal{Q}(X))=\rho tr(X)$, for instance $\mathcal{Q}(X)=H\circ X$ with all diagonal entries being the same and $\mathcal{Q}(X)=UXU^T$ with U being an orthogonal matrix. In practice, Algorithm 3.1 will terminate at iteration $N\ge1$ with $\nu^N\le\varepsilon$. Therefore, the sequence of points $\{(X^k;y^k)\}_{k=1}^N$ generated by Algorithm 3.1 has one of the properties (3.53) and (3.54).*

Proof. Using the same $\mathcal{M}_{X,k}$ and u_k as them in Section 4, together with (3.6), we have

$$u_{min},u_{k-1}>\|\mathcal{Q}\|>\rho \text{ and } X^k=\widehat{X}^{k-1}+\frac{1}{u_{k-1}}\mathcal{Q}(W^k-\widehat{X}^{k-1}) \text{ for all } k=1,2,\cdots,N.$$

In the context of Problem (1.1), we get $tr(W^k)=n$. Choose $\widehat{X}^0=0$ so that $tr(\widehat{X}^0)<n$ then

$$tr(X^k)=\frac{\rho}{u_{k-1}}tr(W^k)+(1-\frac{\rho}{u_{k-1}})tr(\widehat{X}^{k-1})<n \text{ for all } k=1,2,\cdots,N$$

since $\widehat{X}^k$ is one of $X^0, \cdots, X^k$. By setting $c = \max\{tr(\widehat{X}^0), \cdots, tr(\widehat{X}^{N-1})\} < n$, we have

$$\begin{aligned} tr(X^k) - n &= \frac{\rho}{u_{k-1}} tr(W^k) + (1 - \frac{\rho}{u_{k-1}}) tr(\widehat{X}^{k-1}) - n \\ &\le \frac{n\rho}{u_{k-1}} + (1 - \frac{\rho}{u_{k-1}})c - n \le (n-c)(\frac{\rho}{u_{min}} - 1) < 0 \end{aligned}$$

for all $k = 1, \cdots, N$.

On the other hand, it holds $\sigma I - J\Phi(X^k; y^k)(D; d) = (\sigma - e^T d - \langle X^k, \mathcal{Q}(D)\rangle)I + n(\mathcal{A}^*(d) + \mathcal{Q}(D))$. Therefore, we can choose $(\mathcal{Q}(D); d) = (tI; 0)$ with $t < 0$ such that $0 < \sigma \le t(n-c)(\frac{\rho}{u_{min}} - 1)$, then

$$\begin{aligned} \sigma I - J\Phi(X^k; y^k)(D; d) =& \sigma I + t(n - tr(X^k))I \preceq 0 \text{ for all } k = 1, \cdots, N, \\ \frac{\sigma}{\sqrt{\|D\|^2 + \|d\|^2}} \le& \frac{n-c}{\sqrt{n}}(1 - \frac{\rho}{u_{min}})\|\mathcal{Q}\|, \end{aligned}$$

where the second inequality comes from $\|\mathcal{Q}(D)\| \le \|\mathcal{Q}\|\|D\|$. The proof is completed. □

3.7. **Updating eigenvalue tolerance δ^k.** Only if the convergence behavior of Algorithm 3.1 is taken into account, we see that one can choose $\{\delta^k\}$ as any arbitrary convergent decreasing sequence. However, the shortcoming of the sequence, which decreases too quickly, is pointed out in [34]. It says that such choice would result in wasted work computing precise eigenvalues. To avoid this pitfall, we would like to present a choice rule based on the following result.

Theorem 3.6. *Suppose that $\delta^k \downarrow \delta > 0$ as $k \to \infty$ in the context of Proposition 3.5, then*

$$\begin{aligned} &\overline{\lim}_{k\to\infty} u_k l_d^k \le \frac{3 - m_L}{1 - m_L}\delta, \\ &\overline{\lim}_{k\to\infty} \eta^k \le 2\delta. \end{aligned} \tag{3.55}$$

Proof. The second inequality is exactly (3.47).

By the assumption, for $k > N$, it holds $f_l(\widehat{X}^k; \hat{y}^k) - f_u(X^{k+1}; y^{k+1}) < m_L \vartheta^k$, whence

$$\begin{aligned} f_l(\widehat{X}^k; \hat{y}^k) - f_{W^{k+1}}(X^{k+1}; y^{k+1}) <& \frac{1}{1 - m_L}(f_u(X^{k+1}; y^{k+1}) - f_{W^{k+1}}(X^{k+1}; y^{k+1})) \\ =& \frac{1}{1 - m_L}\eta^k. \end{aligned} \tag{3.56}$$

Furthermore, in the light of (3.15), we have $u_k l_d^k = a\hat{\epsilon}^k + f_l(\widehat{X}^k; \hat{y}^k) - f_{W^{k+1}}(X^{k+1}; y^{k+1}) - \tau^k$. Therefore, putting together (3.45), (3.56) and (3.47), we get $u_k l_d^k \le \delta^{k-1} + \frac{1}{1-m_L}\eta^k$, whence (3.55) satisfies. □

The above result guides us to decrease δ^k when $u_k l_d^k$ or η^k is close to its theoretical limit [34], say within 5%. We give the following rule, which is analogous to the one proposed in [34].

At step 5, set $\delta^{k+1} = \min\{10^{-3}|\hat{\lambda}_{ap}^{k+1}|, \delta^k\}$.

At step 6, set $\delta^{k+1} = \begin{cases} \frac{1-m_L}{6-2m_L}\frac{u_k l_d^k}{1.05} & \text{if } u_k l_d^k < \frac{1.05(3-m_L)}{1-m_L}\delta^k, \\ \delta^k & \text{otherwise.} \end{cases}$

4. Numerical experiments

For the sake of validating our approach, we shall now report on computational experiments of our proposed method. We write a MATLAB implementation of Algorithm 3.1, and evaluate its performance on a battery of test problems that we consider below. All the executions are carried out using MATLAB version 7.6 on a 2.10GHz Core 2 laptop computer with 2GB of RAM.

In the numerical experiments, our test problems are similar to those in [49, 50]. They are arising from the nearest correlation matrix problem.

E1: $\mathcal{Q}(X) = HXH$ with $H \succ 0$ generated randomly as follows: `[Q, R]=qr(randn(n)); beta=10∧(-4/(n-1)); H=Q*diag(beta.∧[0:n-1])* Q′`. The matrix C is generated as follows: `T=[ones(n/2), zeros(n/2); zeros(n/2), eye(n/2) ]; -B=-T-1e4*diag(2* rand(n,1)-ones(n,1)); C=H*B*H`.

E2: $\mathcal{Q}(X) = H \circ X$. The matrix H is generated randomly as follows: `tmp=rand(n); H=0.5*(tmp+tmp′)/100`. We generate the matrix C as follows: `x=10.∧[-4:4/(n-1):0]; B=gallery(‘randcorr‘, n* x/sum(x)); tmp=rand(n); tmp=(tmp+tmp′)/2; E=1/norm(tmp,‘fro‘)*tmp; G=B+1e-4*E; C=-H.*G`.

We use two algorithms to solve each test problem, namely,

S1: An iterative solver -Algorithm IP-QSDP [49] that uses PSQMR with constrained preconditioning.

S2: Algorithm 3.1.

In our implementation, we choose the following metric which possesses the desired properties, namely, (1.3), (1.4) and **(A3)**: $(\mathcal{M}_{X,k}, \mathbf{M}_{y,k}) = (\mathcal{I} - \frac{1}{u_k}\mathcal{Q}, I)$ where $u_k > \|\mathcal{Q}\|$ for $k = 0, 1, \cdots$. Such choice results in $\mathcal{T}_k = u_k\mathcal{I}$ and $c_k = 1$.

Note that for any $X \in \mathcal{S}^n$, $\langle X, \mathcal{M}_{X,k}(X)\rangle - \langle X, X\rangle = -\langle \mathcal{Q}(X), X\rangle/u_k \le 0$, whence $\|X\|_{\mathcal{M}_{X,k}} \le \|X\|$ for any $X \in \mathcal{S}^n$ and $k \ge 0$. By virtue of this property, we can get much tighter bound of $(var)_k$.

We reformulate the variation as $(var)_k = f(\widehat{X}^k; \hat{y}^k) - \min\left\{ f(X;y) \middle| (X;y) \in \mathscr{B}_{\mathcal{M}}(\widehat{X}^k; \hat{y}^k, 1)\right\}$.

Using Corollary 3.1 and (3.8), we obtain

$$
\begin{aligned}
&f(\widehat{X}^k; \hat{y}^k) - (var)_k = \min\left\{ f(X;y) \middle| (X;y) \in \mathscr{B}_{\mathcal{M}}(\widehat{X}^k; \hat{y}^k, 1)\right\} \\
&\ge \min\left\{ f(\widehat{X}^k; \hat{y}^k) + \langle g_X^k, X - \widehat{X}^k\rangle + \langle g_y^k, y - \hat{y}^k\rangle - \tau^k \middle| (X;y) \in \mathscr{B}_{\mathcal{M}}(\widehat{X}^k; \hat{y}^k, 1)\right\} \\
&\ge f(\widehat{X}^k; \hat{y}^k) - (u_k(\|\mathcal{M}_{X,k}(d_X^k)\| + \|d_y^k\|) + \tau^k),
\end{aligned}
$$

whence $0 \le (var)_k \le u_k(\|\mathcal{M}_{X,k}(d_X^k)\| + \|d_y^k\|) + \tau^k$.

Hence we replace (3.31) with

$$
e_v^k > \begin{cases} \max\{u_k(\|\mathcal{M}_{X,k}(d_X^k)\| + \|d_y^k\|) + \tau^k, 10(\vartheta^k + a\hat{\epsilon}^k)\} & \text{if } l_d^k \le 1, \\ \min\{u_k(\|\mathcal{M}_{X,k}(d_X^k)\| + \|d_y^k\|) + \tau^k, \vartheta^k + a\hat{\epsilon}^k\} & \text{if } l_d^k > 1. \end{cases}
$$

Accordingly, we replace (3.32), (3.33) and (3.34) with

$$
\theta_v^{k+1} = \begin{cases} \max\{\theta_v^k, 2(\vartheta^k + a\hat{\epsilon}^k)\} & \text{if } l_d^k \le 1, \\ \max\{\theta_v^k, \vartheta^k + a\hat{\epsilon}^k\} & \text{if } l_d^k > 1, \end{cases}
$$

$$
\theta_v^{k+1} = \begin{cases} \min\{\theta_v^k,\ u_k(\|\mathcal{M}_{X,k}(d_X^k)\| + \|d_y^k\|) + \tau^k\} & \text{if } l_d^k \le 1, \\ \min\{\theta_v^k, \vartheta^k + a\hat{\epsilon}^k, u_k(\|\mathcal{M}_{X,k}(d_X^k)\| + \|d_y^k\|) + \tau^k\} & \text{if } l_d^k > 1, \end{cases}
$$

$$
e_v^k > \begin{cases} \max\{\theta_v^{k+1}, 10(\vartheta^k + a\hat{\epsilon}^k)\} & \text{if } l_d^k \le 1, \\ \min\{\theta_v^{k+1}, \vartheta^k + a\hat{\epsilon}^k\} & \text{if } l_d^k > 1, \end{cases}
$$

respectively. Note that the above replacements do not affect the convergence of the proposed algorithm.

For time saving, the stopping criterion in Algorithm 3.1 is replaced by

$$\tilde{\nu}^k = \max\{l_d^k, \min\{\eta^k, \tau^k\}\} \leq \varepsilon. \tag{4.1}$$

Thus (3.53) and (3.54) are reduced to

$$d((\widehat{X}^k; \hat{y}^k), \mathcal{O}_*) \leq \frac{\varepsilon}{\frac{\sigma}{\sqrt{\|D\|^2+\|d\|^2}} - u_{max}\sqrt{\varepsilon}} \text{ and } d((X^{k+1}; y^{k+1}), \mathcal{O}_*) \leq \frac{\varepsilon}{\frac{\sigma}{\sqrt{\|D\|^2+\|d\|^2}} - u_{max}\sqrt{\varepsilon}},$$

respectively, which indicate that (4.1) is reasonable.

The dominant process in Algorithm 3.1 is the eigenvalue estimation. Considering that eigenvalue problems in question are not necessarily sparse, we adopt the existing program **eigifp** [20] to solve them, whose underlying algorithm is an inverse free preconditioned Krylov subspace projection method [11].

In Step 2 of Algorithm 3.1, we need to compute the subproblem (SQSDP). With the help of the symmetrized Kronecker product [1], using the analogous technique as that presented in [14], we can cast (SQSDP) as a quadratic semidefinite programming with small size. We solve the quadratic semidefinite programming by a feasible primal–dual interior– point method, whose essential structure is that of Mehrotra-type predictor-corrector primal-dual path-following algorithm in [50]. The method used here differs in that we should guarantee the feasibility at each iteration. In addition, the Nesterov-Todd symmetrization scheme [48] is employed.

The initial iterate for Algorithm 3.1 is given by $X^0 = 0,\ y^0 = 0$. The value of parameters present in the algorithm are listed as follows:

$$\begin{aligned}
&(\text{the upper bound of the number of columns of all } P_k s) \leq 30,\\
&m_L = 0.1, m_R = 0.55, \delta^0 = 10^{-6}, \theta_v^0 = 10^{15},\\
&u_{min} = \begin{cases} \|\mathcal{Q}\| + 0.01 & \text{if } \|\mathcal{Q}\| > 1, \\ 1.01\|\mathcal{Q}\| & \text{otherwise.} \end{cases}, u_0 = \max\left\{u_{min}, \sqrt{\|\mathcal{A}(\overline{W}_0) - b\|^2 + \|\mathcal{Q}(\overline{W}_0)\|^2}\right\},
\end{aligned}$$

where $\overline{W}_0 = a\hat{v}_{ap}^0(\hat{v}_{ap}^0)^T$. The stopping criterion is based on (4.1) with

$$\varepsilon = \begin{cases} 10^{-4} & \text{test problems in } \texttt{E1}, \\ 10^{-3} & \text{test problems in } \texttt{E2} \text{ and } n \leq 1600, \\ 3 \cdot 10^{-3} & \text{test problems in } \texttt{E2} \text{ and } n = 2000, \end{cases}$$

in the implementation.

The performance results are given in the following Table 1, where we compare the proposed inexact spectral bundle method S2 to Toh's method S1. The columns corresponding to "iter" give the total number of iterations. The column corresponding to "value(p)" stands for the minimum value of the objective function in Problem (1.1) computed by S1, while the column corresponding to "value(d)" signifies the opposite number of the minimum value of the objective function in Problem (2.1) solved by S2. Note that both of the two quantities discard the constant $\langle B, \mathcal{L}^2(B)\rangle/2$. The columns corresponding to "ϕ" and "ν" present the accuracy measure defined by (46) in [49] and (4.1), respectively. The column "SS" gives the number of serious steps, whereas the columns "NgX" and "Ngy" refer to the norms of $\mathcal{Q}(X - W)$ and $b - \mathcal{A}(W)$, respectively, which give rise to the approximate subgradient resulting from the optimal solution of the last quadratic semidefinite subproblem. We calculate elapsed time in seconds. Those entries being "err" mean that the algorithm terminates with out-of-memory errors.

Table 1 provides some information that we sum up below.

Table 1. Performance of Algorithm S1 and S2 on the problems sets E1 and E2

		S1				S2						
	n	iter	time(s)	value(p)	ϕ	iter	SS	time(s)	value(d)	NgX	Ngy	ν
E1	200	17	40.2	-21.647	3.16e-7	72	25	77	-21.919	0.00599	0.0285	6.84e-5
	400	21	338.4	-48.745	1.88e-7	64	24	179.3	-49.681	0.0116	0.0975	9.70e-5
	800	12	661.4	-97.598	9.48e-1	62	23	605.7	-109.13	0.0226	0.137	9.69e-5
	1600	13	4638.8	-191.60	6.87e-1	82	31	4368.7	-214.21	0.0377	0.194	9.75e-5
	2000	err	err	err	err	102	36	10532.2	-340.26	0.0505	0.0849	7.81e-5
	3000	err	err	err	err	130	47	39733.5	-477.73	0.0684	0.072	5.26e-5
E2	200	14	17.9	-2.3209	6.51e-7	238	37	624.6	-3.3959	0.388	0.474	8.53e-4
	400	14	109	-4.6572	7.40e-7	126	55	1119.6	-10.065	0.614	0.918	9.90e-4
	800	15	580.7	-9.2369	1.90e-7	119	27	1974.5	-27.284	1.09	1.75	8.53e-4
	1600	11	1539.5	-18.441	1.71e-4	88	28	5505.2	-62.884	1.87	1.28	9.93e-4
	2000	err	err	err	err	164	25	14122	-82.826	2.16	2.17	2.95e-3

(1) The number of serious steps required by our proposed Algorithm S2 changes slowly but steadily with the problem dimension n. In all the test problems, the number of such iterations is less than 60.

(2) As a whole, Algorithm S2 outperforms Algorithm S1 on the problem set E1. We see that our method can find reasonably accurate solution in E1-400 in about half the time required by Algorithm S1. In addition, Algorithm S1 stops since it hits the maximum of interior-point iterations allowed or the gap is too large based on its criterion in E1-800 and E1-1600, while we can achieve $(10^{-4})^{F2}$**-optimal** solutions by Algorithm S2 on the same problems. On the other hand, these results are also comparative with those in [46]. Indeed in the terms of CPU time, our proposed algorithm is faster than S1 and slower than the proposed method in [46]. We explain the case when $n \geq 2000$ in the next item.

(3) Algorithm S1 is not able to solve the test problems in E1-2000, E1-3000 and E2-2000 since interior-point methods are not applicable because of memory requirements, while Algorithm S2 performs well on these problems. This demonstrates that it is hopeful for bundle methods to solve large-scale problems with the size of the matrix variables more than 2000. However, it also displays that there is little hope for bundle methods to end up within reasonable time.

(4) Algorithm S1 outperforms Algorithm S2 on the problem set E2 with $n \leq 1600$. It should be clear from the examples of Table 1 that Algorithm S2 spends much more time but obtains less accurate solutions than Algorithm S1. Furthermore, the rather large norm of the subgradient of the last three problems shows that the values cannot be expected to be "good" approximate of the objective function, which also is illustrated by comparing the column corresponding to "value(d)" with the one corresponding to "value(p)" with respect to E2. We will improve the performance of our algorithm in the case when $\mathcal{Q}(X) = H \circ X$ in the future.

(5) From the column corresponding to "time(s)" with respect to S2, we see that the computational time of E1 is much less than that of E2. This arises from the fact that $\mathcal{Q}(X) = HXH$ with $H \succ 0$ can be viewed as a special case of $\mathcal{Q}(X) = H \circ X$. Indeed, using the same technique as its used in linear semidefinite programs [13, Proposition 2.1.3], we can transform E1 into E2 by scaling the variables and operators. Recalling that, in E1, $H = QDiag(\beta)Q^T$ with $\boldsymbol{\beta}$=beta.∧[0:n-1], we

scale $(X, y, Z, W, C, \mathcal{A}, \mathcal{Q})$ to $(\widetilde{X}, y, \widetilde{Z}, \widetilde{W}, \widetilde{C}, \widetilde{\mathcal{A}}, \widetilde{\mathcal{Q}})$ as follows:

$$\widetilde{X} = Q^T X Q,\ \ \widetilde{Z} = Q^T Z Q,\ \ \widetilde{W} = Q^T W Q, \widetilde{C} = Q^T C Q,$$

$$\widetilde{\mathcal{A}}(\widetilde{\cdot}) = (\langle Q^T A_1 Q, \widetilde{\cdot}\rangle, \cdots, \langle Q^T A_m Q, \widetilde{\cdot}\rangle)^T, \widetilde{\mathcal{Q}}(\widetilde{\cdot}) = Q^T H Q(\widetilde{\cdot}) Q^T H Q,$$

then it holds $\langle C, X\rangle = \langle \widetilde{C}, \widetilde{X}\rangle, \mathcal{A}(X) = \widetilde{\mathcal{A}}(\widetilde{X}), \langle X, \mathcal{Q}(X)\rangle = \langle \widetilde{X}, \widetilde{\mathcal{Q}}(\widetilde{X})\rangle$. Hence, such transformation will get the same convex quadratic semidefinite programming in terms of $(\widetilde{X}, y, \widetilde{Z})$ and $\widetilde{\mathcal{Q}}(\widetilde{X}) = \widetilde{H} \circ \widetilde{X}$, where $\widetilde{H} = \beta\beta^T$.

5. Conclusion

We have proposed an inexact spectral bundle method for solving linearly constrained convex quadratic semidefinite programming. Convergence analysis of the proposed algorithm has been discussed. In the context of inexact computation, it suffices to estimate $f(X; y)$ and its subgradients sequentially with a specified sequence of tolerance δ^k. The proposed algorithm finds two sequences $\{(X^k; y^k)\}$ and $\{W^k\}$ simultaneously, whose limit points are $(\frac{2\delta}{1-m_L})^{F1}$**-optimal** solutions for (EigForm) and (P) respectively, when $\delta^k \downarrow \delta > 0$. In addition, a global error bound is given under a Slater type condition, which also demonstrates the practicality of the stopping rule in our proposed method.

The proposed method does require first-order information and it is easy to implement. Numerical experiments on a set of the nearest correlation matrix problem with matrices of order up to 3000 show our method is efficient. However, it should be clear from our numerical results that, in terms of CPU time, the performance of our method on `E2` is much slower than its performance on `E1`. This may be due to the high density of the problem data. Therefore, one important future work is to overcome this drawback, possibly by using a parallel implementation, see, e.g. [19]. On the other hand, it is of interest to ask whether our proposed method can be extended to solve the convex quadratic symmetric cone programming.

Acknowledgements

I am very grateful to my supervisor Chua Chek Beng for his helpful suggestions and comments. The author is supported by the research scholarship provided by the Nanyang Technological University, Singapore.

References

[1] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton. Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numrical results. *SIAM J. Optim.*, 8(3):746–768, 1998.

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

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

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

[6] 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.

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

[8] R. Fletecher. Semi-definite matrix constraints in optimization. *SIAM J. Control. Optim.*, 23:493–523, 1985.

[9] T. Fushiki. Estimation of positive semidefinite correlation matrices by using convex quadratic semidefinite programming. *Neural Comput.*, 21 (7):2028–2048, 2009.

[10] Y. Gao and D. Sun. Calibrating least squares covariance matrix problems with equality and inequality constraints. *SIAM J. Matrix Anal. Appl.*, 31:1432–1457, 2009.

[11] 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.

[12] A. Graham. *Kronecker products and matrix calculus with applications.* Ellis Horwood Limited, Chichester, England, 1981.

[13] C. Helmberg. Semidefinite programming for combinatorial optimization. *ZIB-Report ZR-00-34, Konrad-Zuse-Zentrum Berlin*, October 2000.

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

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

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

[17] J. B. Hiriart-Urruty and D.Ye. Sensitivity analysis of all eigenvalue of a symmetric matrix. *Numer. Math.*, 70(1):45–72, 1995.

[18] J. B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithm II, volume 306 of Grundlehren der mathematischen Wissenschaften.* Springer, Berlin, Heidelberg, 1993.

[19] I. D. Ivanov and E. D. Klerk. Parallel implementation of a semidefinite programming solver based on CSDP on a distributed memory cluster. *Optim. Methods and Soft.*, 25(30):405–420, 2010.

[20] 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.

[21] S. Jie and P. Li-Ping. A quasi-newton bundle method based on approximate subgradients. *J. Appl. Math. Comput.*, 23 (1-2):361–367, 2007.

[22] A. Jourani and J. J.Ye. Error bounds for eigenvalue and semidefinite matrix inequality systems. *Math. Programming. Ser. B*, 104:525–540, 2005.

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

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

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

[26] M. Kojima, S. Shindoh, and S. Hara. Interior-point methods for the monotone linear complementarity problem in symmetric matrices. *SIAM J. Optim.*, 7:86–125, 1997.

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

[28] A.S. Lewis. Nonsmooth analysis of eigenvalues. *Math. Programming*, 84:1–24, 1999.

[29] L. Li and K.C. Toh. A polynomial-time inexact primal-dual infeasible path-following algorithm for convex quadratic SDP. *Pacific J. Optim.*, Oct. 2009. accepted.

[30] P. Loridan. Necessary conditions for ϵ-optimality. *Math. Programming Study*, 19:140–152, 1982.

[31] M. M. Mäkelä. Survey of bundle methods for nonsmooth optimization. *Optim. Methods and Soft.*, 17(1):1–29, 2002.

[32] M. M. Mäkelä and P. Neittaanmäki. *Nonsmooth Optimization.* World Scientific, Singapore, 1992.

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

[34] S. A. Miller and R. S. Smith. A bundle method for efficiently solving large structured linear matrix inequalities. In *Proc. Amer. Contr. Conf.*, pages 1405–1409, Chicago, IL, June 2000.

[35] B. S. Mordukhovich and Y. Shao. Nonsmooth sequential analysis in asplund space. *Trans. Amer. Math. Soc.*, 348:215–220, 1996.

[36] T. T. V. Nguyen, J. J. Strodiot, and V. H. Nguyen. A bundle method for solving equilibrium problems. *Math. Programming, Ser. B*, 116(1-2):529–552, 2009.

[37] J. W. Nie and Y. X. Yuan. A predictor-corrector algorithm for QSDP combining Dikin-type and Newton centering steps. *Ann. Oper. Res.*, 103:115–133, 2001.

[38] F. Oustry. A second-order bundle method to minimize the maximum eigenvale function. *Math. Programming. Ser. A.*, 89:1–33, 2000.

[39] M. L. Overton. Large-scale optimization of eigenvalues. *SIAM J. Optim.*, 2:88–120, 1992.
[40] H.-D. Qi and D. Sun. An augmented lagrangian dual approach for the H-weighted nearest correlation matrix problem. *IMA J. Numer. Anal.*, in press, doi:10.1093/imanum/drp031, 2010.
[41] 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.
[42] M. Rotunno and R.A. de Callafon. A bundle method for solving the fixed order control problem. In *Proc. 41st IEEE Conference on Decision and Control*, volume 3, pages 3156–3161, Las Vegas, NV, USA, 2002.
[43] C. Sagastizábal and M. Solodov. On the relation between bundle methods for maximal monotone inclusions and hybrid proximal point algorithms. *Stud. Comput. Math.*, 8:441–455, 2001.
[44] 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.
[45] M. V. Solodov. A bundle method for a class of bilevel nonsmooth convex minimization problems. *SIAM J. Optm.*, 18(1):242–259, 2007.
[46] J. Sun and S. Zhang. A modified alternating direction method for convex quadratically constrained quadratic semidefinite programs. *Eur. J. Oper. Research*, 207(3):1210–1220, 2010.
[47] C. H. Teo, S. V. N. Vishwanathan, A. Smola, and Q.V. Le. Bundle methods for regularized risk minimization. *J. Machine Learning Research*, 11:311–365, 2010.
[48] M. J. Todd, K. C. Toh, and R. H. Tütüncü. On the Nesterov-Todd direction in semidefinite programming. *SIAM J. Optim.*, 8:769–796, 1998.
[49] K. C. Toh. An inexact primal-dual path-following algorithm for convex quadratic SDP. *Math. Programming*, 112:221–254, 2008.
[50] K. C. Toh, R. H. Tütüncü, and M. J. Todd. Inexact primal-dual path-following algorithms for a special class of convex quadratic SDP and related problems. *Pacific J. Optim. (special issue dedicated to Masakazu Kojima's 60th birthday)*, 3:135–164, 2007.
[51] K.C. Toh. User guide for QSDP-0– a MATLAB software package for convex quadratic semidefinite programming. *Working paper, Department of Mathematics, National University of Singapore, Singapore*, February 2010. http://www.math.nus.edu.sg/~mattohkc/QSDP-guide.pdf.
[52] Z. Wu and J. J. Ye. First-order and second-order conditions for error bounds. *SIAM J. Optim.*, 14:621–645, 2003.
[53] X.Y. Zhao. *A semismooth Newton-CG augmented Lagrangian method for large scale linear and convex quadratic SDPs*. PhD thesis, National University of Singapore, 2009.

Appendix A. Proofs of Proposition 3.5 and 3.6

Proof of Proposition 3.5:

Proof. The proof hinges on the following two observations:

$$\Delta_k := f(X^k; y^k) - f_{W^k}(X^k; y^k) \to 0, \tag{A.1}$$

$$b - \mathcal{A}(W^k) \to 0 \text{ and } \mathcal{Q}(X^k - W^k) \to 0 \tag{A.2}$$

as $k \to \infty$, we will show them later.

First we prove that the sequence $\{L^{k-1}(X^k, y^k, W^k)\}_{k=N}^{\infty}$ is nondecreasing.

For any $k \geq N$, it is easy to see that

$$L^k(X, y, W) = f_W(X; y) + \frac{u_k}{2}(\|X - \widehat{X}\|^2_{\mathcal{M}_{X,k}} + \|y - \hat{y}\|^2_{\mathbf{M}_{y,k}}).$$

By Proposition 3.2,

$$W^{k+1} = arg \max_{W \in \widehat{\mathcal{W}}^k} L^k(X^{k+1}, y^{k+1}, W) = arg \max_{W \in \widehat{\mathcal{W}}^k} f_W(X^{k+1}; y^{k+1}). \tag{A.3}$$

If there is no descent step from iteration N, by step 5, $u_{k+1} \geq u_k$ for any $k \geq N$.

Using (3.20), (1.4) and $u_k \geq u_{k-1}$, we have

$$
\begin{aligned}
&L^k(X^{k+1}, y^{k+1}, W^k)\\
=&L^k(X^N_{min}(W^k), y^N_{min}(W^k), W^k) + \frac{1}{2}\|X^{k+1} - X^N_{min}(W^k)\|^2_{\mathcal{T}_k} + \frac{u_k}{2}\|y^{k+1} - y^N_{min}(W^k)\|^2_{\mathbf{M}_{y,k}}\\
\geq &L^k(X^k, y^k, W^k) + \frac{u_k}{2}(\|X^{k+1} - X^k\|^2_{\mathcal{M}_{X,k}} + \|y^{k+1} - y^k\|^2_{\mathbf{M}_{y,k}})\\
\geq &L^{k-1}(X^k, y^k, W^k) + \frac{u_{min}}{2}(\|X^{k+1} - X^k\|^2_{\mathcal{M}_{X,k}} + \|y^{k+1} - y^k\|^2_{\mathbf{M}_{y,k}}) \geq L^{k-1}(X^k, y^k, W^k).
\end{aligned}
$$

Thus for any $k \geq N$,

$$
L^{k-1}(X^k, y^k, W^k) \leq L^k(X^{k+1}, y^{k+1}, W^k) \leq L^k(X^{k+1}, y^{k+1}, W^{k+1}) \leq f(\widehat{X}; \hat{y}), \tag{A.4}
$$

where the middle inequality follows from (A.3) and $W^k \in \widehat{\mathcal{W}^k}$ by step 7, and the last inequality follows from (3.21). Therefore, there exists some $f_* \in \mathbb{R}$ such that

$$
L^{k-1}(X^k, y^k, W^k) \uparrow f_* \leq f(\widehat{X}; \hat{y}) \text{ as } k \to \infty, \tag{A.5}
$$

whence $f_* \leq f_* + \lim\limits_{k\to\infty} \frac{u_{min}}{2}(\|X^{k+1} - X^k\|^2_{\mathcal{M}_{X,k}} + \|y^{k+1} - y^k\|^2_{\mathbf{M}_{y,k}}) \leq f_*$. Therefore

$$
\|X^{k+1} - X^k\|_{\mathcal{M}_{X,k}} \to 0 \text{ and } \|y^{k+1} - y^k\|_{\mathbf{M}_{y,k}} \to 0 \text{ as } k \to \infty,
$$

whence

$$
\|X^{k+1} - X^k\| \to 0 \text{ and } \|y^{k+1} - y^k\| \to 0 \text{ as } k \to \infty. \tag{A.6}
$$

On the one hand, set $(g^{k-1}_{X\_ap}; g^{k-1}_{y\_ap}) = (\mathcal{Q}(X^k - W^k_{ap}); b - \mathcal{A}(W^k_{ap}))$, then, by step 3, Proposition 2.4 and (2.4), we have $(g^{k-1}_{X\_ap}; g^{k-1}_{y\_ap}) \in \partial_{\epsilon^k} f(X^k; y^k)$.

Define a linear function on $\mathcal{S}^n \times \mathbb{R}^m$ by

$$
l^k(X; y) = f_l(X^k; y^k) + \langle g^{k-1}_{X\_ap}, X - X^k\rangle + \langle g^{k-1}_{y\_ap}, y - y^k\rangle,
$$

since $W^k_{ap} \in \widehat{\mathcal{W}^k}$ and (3.2), we have

$$
\begin{aligned}
l^k(X; y) =&\langle -C - \mathcal{A}^*(y) - \mathcal{Q}(X), W^k_{ap}\rangle + b^T y + \frac{1}{2}\langle X^k, \mathcal{Q}(X^k)\rangle + \langle \mathcal{Q}(X^k), X - X^k\rangle\\
\leq &\langle -C - \mathcal{A}^*(y) - \mathcal{Q}(X), W^k_{ap}\rangle + b^T y + \frac{1}{2}\langle X, \mathcal{Q}(X)\rangle \leq f_{\widehat{\mathcal{W}^k}}(X; y).
\end{aligned}
$$

Therefore, using the expressions of $L^k(X, y, W)$ and $l^k(X, y)$, it implies

$$
\begin{aligned}
0 \leq &\Delta_k \leq f_u(X^k; y^k) - f_{W^k}(X^k; y^k) = l^k(X^k; y^k) + a\epsilon^k - f_{W^k}(X^k; y^k)\\
&\leq l^k(X^{k+1}; y^{k+1}) - f_{W^k}(X^k; y^k) - \langle g^{k-1}_{X\_ap}, X^{k+1} - X^k\rangle - \langle g^{k-1}_{y\_ap}, y^{k+1} - y^k\rangle + a\epsilon^k\\
&\leq f_{W^{k+1}}(X^{k+1}; y^{k+1}) - f_{W^k}(X^k; y^k) - \langle g^{k-1}_{X\_ap}, X^{k+1} - X^k\rangle - \langle g^{k-1}_{y\_ap}, y^{k+1} - y^k\rangle + \delta^{k-1}\\
&\leq L^k(X^{k+1}, y^{k+1}, W^{k+1}) - L^{k-1}(X^k, y^k, W^k) - \frac{u_k}{2}(\|X^{k+1} - \widehat{X}\|^2_{\mathcal{M}_{X,k}} + \|y^{k+1} - \hat{y}\|^2_{\mathbf{M}_{y,k}})\\
&\quad + \frac{u_{k-1}}{2}(\|X^k - \widehat{X}\|^2_{\mathcal{M}_{X,k-1}} + \|y^k - \hat{y}\|^2_{\mathbf{M}_{y,k-1}}) + \|b - \mathcal{A}(W^k_{ap})\|\|y^k - y^{k+1}\|\\
&\quad + \|X^{k+1} - X^k\|\|\mathcal{Q}(W^k_{ap} - X^k)\| + \delta^{k-1}.
\end{aligned}
$$

Since $W^k, W^k_{ap} \in a\mathcal{W}$ for any $k \geq N$, $\{\|X^k\|\}$, $\{\|y^k\|\}$, $\{\|W^k\|\}$ and $\{\|W^k_{ap}\|\}$are bounded.

In what follows, we show that both (A.1) and (A.2) are valid.

Case 1. $u_k \uparrow \infty$ as $k \to \infty$.

By (3.35) and (3.36), we get

$$X^k - \widehat{X} = \frac{\mathcal{M}_{X,k-1}^{-1}(\mathcal{Q}(W^k - X^k))}{u_{k-1}} \to 0,\ \ y^k - \hat{y} = \frac{\mathbf{M}_{y,k-1}^{-1}(\mathcal{A}(W^k) - b)}{u_{k-1}} \to 0 \tag{A.7}$$

as $k \to \infty$ since both $\{\|W^k\|\}$ and $\{\|X^k\|\}$ are bounded. Moreover,

$$\begin{aligned}&\frac{u_{k-1}}{2}(\|X^k - \widehat{X}\|^2_{\mathcal{M}_{X,k-1}} + \|y^k - \hat{y}\|^2_{\mathbf{M}_{y,k-1}}) - \frac{u_k}{2}(\|X^{k+1} - \widehat{X}\|^2_{\mathcal{M}_{X,k}} + \|y^{k+1} - \hat{y}\|^2_{\mathbf{M}_{y,k}})\\ \le&\frac{1}{2u_{k-1}}(\|\mathcal{Q}(W^k - X^k)\|^2_{\mathcal{M}^{-1}_{X,k-1}} + \|\mathcal{A}(W^k) - b\|^2_{\mathbf{M}^{-1}_{y,k-1}}),\end{aligned}$$

thus

$$\begin{aligned}\Delta_k \le& L^k(X^{k+1}, y^{k+1}, W^{k+1}) - L^{k-1}(X^k, y^k, W^k) + \frac{1}{2u_{k-1}}(\|\mathcal{Q}(W^k - X^k)\|^2_{\mathcal{M}^{-1}_{X,k-1}}\\ &+ \|\mathcal{A}(W^k) - b\|^2_{\mathbf{M}^{-1}_{y,k-1}}) + \|b - \mathcal{A}(W^k_{ap})\|\|y^k - y^{k+1}\| + \|X^{k+1} - X^k\|\|\mathcal{Q}(W^k_{ap} - X^k)\| + \delta^{k-1},\end{aligned}$$

Therefore, together with (A.5), (A.6) and $\delta^k \to 0$, it yields $\lim_{k\to\infty} \eta^{k-1} = 0$ and $\lim_{k\to\infty} \Delta_k = 0$.

Next, we will show that $e_v^k \to 0$ as $k \to \infty$.

For any $k > N$, on the one hand,

$$\begin{aligned}e_v^{k-1} =& f_u(\widehat{X}^{k-1}; \hat{y}^{k-1}) - f_{W^k_{ap}}(\widehat{X}^{k-1}; \hat{y}^{k-1})\\ \ge& a\lambda_1(-C - \mathcal{A}^*(\hat{y}) - \mathcal{Q}(\widehat{X})) - \langle -C - \mathcal{A}^*(\hat{y}) - \mathcal{Q}(\widehat{X}), W^k_{ap}\rangle \ge 0.\end{aligned}$$

We recall that $a\hat{\epsilon}^k \le \min\{\delta^{k-1}, 10\delta^k\}$, hence $\hat{\epsilon}^k \to 0$ if $\delta^k \to 0$.

On the other hand, it is known that $\lambda_1(\cdot)$ is locally Lipschitz continuous, then by step 4, combining with (A.7) and $\delta^k \to 0$, we have

$$\begin{aligned}e_v^{k-1} \le& a\lambda_1(-C - \mathcal{A}^*(\hat{y}) - \mathcal{Q}(\widehat{X})) - \langle -C - \mathcal{A}^*(\hat{y}) - \mathcal{Q}(\widehat{X}), W^k_{ap}\rangle + a\hat{\epsilon}^{k-1}\\ \le& a\lambda_1(-C - \mathcal{A}^*(\hat{y}) - \mathcal{Q}(\widehat{X})) - a\lambda_1(-C - \mathcal{A}^*(y^k) - \mathcal{Q}(X^k)) + a\epsilon^k + a\hat{\epsilon}^{k-1}\\ &- \langle y^k - \hat{y}^k, \mathcal{A}(W^k_{ap})\rangle - \langle \mathcal{Q}(X^k - \widehat{X}), W^k_{ap}\rangle\\ \le& a((L\|\mathcal{A}^{\mathcal{T}}\| + \|\mathcal{A}(W^k_{ap})\|)\|y^k - \hat{y}\| + (L\|\mathcal{Q}\| + \|\mathcal{Q}(\mathcal{A}(W^k_{ap}))\|)\|X^k - \widehat{X}\| + \epsilon^k + \hat{\epsilon}^{k-1})\\ \to& 0 \text{ as } k \to \infty,\end{aligned}$$

where L is the Lipschitz constant.

By step 6, the above expression yields to $\theta_v^k \to 0$, therefore $\vartheta^{k-1} + a\hat{\epsilon}^{k-1} \to 0$, and then by (3.29), it holds $u_{k-1}l_d^{k-1} + \tau^{k-1} \to 0$ as $k \to \infty$, whence $b - \mathcal{A}(W^k) = -u_{k-1}\mathbf{M}_{y,k-1}(d_y^{k-1}) \to 0$ and $\mathcal{Q}(X^k - W^k) = -u_{k-1}\mathcal{M}_{X,k-1}(d_X^{k-1}) \to 0$ as $k \to \infty$, from (A.7) and (3.8).

Case 2. $u_k \uparrow \bar{u} \in (0, \infty)$ as $k \to \infty$.

First we prove that $\Delta_k \to 0$ as $k \to \infty$.

Since $u^{k-1} \le u^k \le \bar{u}$ for any $k \ge N$, with the help of (1.4),

$$\begin{aligned}\Delta_k \le& L^k(X^{k+1}, y^{k+1}, W^{k+1}) - L^{k-1}(X^k, y^k, W^k) + u_k\langle X^k - X^{k+1}, \mathcal{M}_{X,k}(X^k - \widehat{X})\rangle\\ &+ u_k\langle y^k - y^{k+1}, \mathbf{M}_{y,k}(y^k - \hat{y})\rangle - \frac{u_k}{2}\|X^k - X^{k+1}\|^2_{\mathcal{M}_{X,k}} - \frac{u_k}{2}\|y^k - y^{k+1}\|^2_{\mathbf{M}_{y,k}}\\ &+ \|b - \mathcal{A}(W^k_{ap})\|\|y^k - y^{k+1}\| + \|X^{k+1} - X^k\|\|\mathcal{Q}(W^k_{ap} - X^k)\| + \delta^{k-1}\\ \le& L^k(X^{k+1}, y^{k+1}, W^{k+1}) - L^{k-1}(X^k, y^k, W^k) + (\bar{u}\|\mathbf{M}_{y,k}(y^k - \hat{y})\| + \|b - \mathcal{A}(W^k_{ap})\|)\|y^k - y^{k+1}\|\\ &+ (\bar{u}\|\mathcal{M}_{X,k}(X^k - \widehat{X})\| + \|\mathcal{Q}(W^k_{ap} - X^k)\|)\|X^{k+1} - X^k\| + \delta^{k-1}.\end{aligned}$$

Therefore, together with (A.5), (A.6), $\lim_{k\to\infty}\delta^k = 0$ and the boundness of $\{\|X^k\|\}$, $\{\|y^k\|\}$ and $\{\|W^k_{ap}\|\}$, it follows $\Delta_k \to 0$ as $k\to\infty$.

In the following, we prove that, in both cases, it holds

$$f_{W^k}(X^k;y^k) \to f(\widehat{X};\hat{y}) \text{ as } k\to\infty. \tag{A.8}$$

If no descent steps occur, then $f_l(\widehat{X}^{k-1};\hat{y}^{k-1}) - f_u(X^k;y^k) < m_L(f_l(\widehat{X}^{k-1};\hat{y}^{k-1}) - f_{W^k}(X^k;y^k))$, thus for $k>N$, it follows

$$\begin{aligned} f_u(X^k;y^k) - f_{W^k}(X^k;y^k) >&(1-m_L)(f_l(\widehat{X}^{k-1};\hat{y}^{k-1}) - f_{W^k}(X^k;y^k)) \\ \geq& -a\hat{\epsilon}^{k-1}(1-m_L) \geq -\delta^{k-2}(1-m_L). \end{aligned}$$

whence (A.8) holds.

On the other hand, $W^k\in a\mathcal{W}$, by (3.20), we obtain, if $u_k \uparrow \bar{u}\in(0,\infty)$ as $k\to\infty$,

$$\begin{aligned} f(\widehat{X};\hat{y}) =& \max_{W\in a\mathcal{W}} f_W(\widehat{X};\hat{y}) \geq f_{W^k}(\widehat{X};\hat{y}) + \frac{u_{k-1}}{2}(\|\widehat{X}-\widehat{X}\|^2_{\mathcal{M}_{X,k-1}} + \|\hat{y}-\hat{y}\|^2_{\mathbf{M}_{y,k-1}}) \\ =&L^{k-1}(\widehat{X},\hat{y},W^k) = L^{k-1}(X^k,y^k,W^k) + \frac{1}{2}\|\widehat{X}-X^k\|^2_{\mathcal{T}_{k-1}} + \frac{u_{k-1}}{2}\|\hat{y}-y^k\|^2_{\mathbf{M}_{y,k-1}} \\ \geq& f_{W^k}(X^k;y^k) + \bar{u}\|\widehat{X}-X^k\|^2_{\mathcal{M}_{X,k-1}} + \bar{u}\|\hat{y}-y^k\|^2_{\mathbf{M}_{y,k-1}}, \end{aligned}$$

namely, $\bar{u}\|\widehat{X}-X^k\|^2_{\mathcal{M}_{X,k-1}} + \bar{u}\|\hat{y}-y^k\|^2_{\mathbf{M}_{y,k-1}} \leq f(\widehat{X};\hat{y}) - f_{W^k}(X^k;y^k)$, whence $X^k\to\widehat{X}$ and $y^k\to\hat{y}$ as $k\to\infty$ by (A.8). Therefore,

$$b-\mathcal{A}(W^k) = u_{k-1}\mathbf{M}_{y,k-1}(\hat{y}-y^k)\to 0,\ \mathcal{Q}(X^k-W^k) = u_{k-1}\mathcal{M}_{X,k-1}(\widehat{X}-X^k)\to 0$$

as $k\to\infty$.

Hence, in both cases, together with $(\mathcal{Q}(X^k-W^k);b-\mathcal{A}(W^k))\in\partial f_{\widehat{\mathcal{W}}^{k-1}}(X^k;y^k)$ by (3.9), we imply that, for any $(X;y)\in\mathcal{S}^n\times\mathbb{R}^m$,

$$f(X;y)\geq f_{\widehat{\mathcal{W}}^{k-1}}(X;y) \geq f_{W^k}(X^k;y^k) + \langle\mathcal{Q}(X^k-W^k),X-X^k\rangle + \langle b-\mathcal{A}(W^k),y-y^k\rangle \to f(\widehat{X};\hat{y})$$

as $k\to\infty$ by (A.8) and (A.2), whence $(\widehat{X};\hat{y})\in arg\min_{X,y} f(X;y)$. Moreover, combining with the boundness of c_k, we obtain

$$c_{k-1}u_{k-1}\sqrt{l_d^{k-1}} = c_{k-1}(\|\mathcal{M}^{-1/2}_{X,k-1}(g_X^{k-1})\|^2 + \|\mathbf{M}^{-1/2}_{y,k-1}g_y^{k-1}\|^2)^{1/2}\to 0 \ as\ k\to\infty.$$

Moreover, $0\leq\min\{\eta^{k-1},\tau^{k-1}\}\leq\eta^{k-1}\to 0$ as $k\to\infty$, hence $\lim_{k\to\infty}\nu^{k-1}=0$. □

Proof of Proposition 3.6:

Proof. By the assumption, for any h, $f(X^h;y^h)\geq f(\widetilde{X};\tilde{y})$. Thus, the inexact computation of $f(X;y)$ yields

$$\begin{aligned} &f_l(X^h;y^h)\geq\langle -C-\mathcal{A}^*(\tilde{y})-\mathcal{Q}(\widetilde{X}),W^{h+1}\rangle + b^T\tilde{y} + \frac{1}{2}\langle\widetilde{X},\mathcal{Q}(\widetilde{X})\rangle - a\epsilon^h \\ =&f_{W^{h+1}}(X^{h+1};y^{h+1}) + \langle b-\mathcal{A}(W^{h+1}),\tilde{y}-y^{h+1}\rangle + \langle\mathcal{Q}(X^{h+1}-\widetilde{X}),W^{h+1}-\widetilde{X}\rangle \\ &-\frac{1}{2}\langle X^{h+1}-\widetilde{X},\mathcal{Q}(X^{h+1}-\widetilde{X})\rangle - a\epsilon^h \end{aligned}$$

By virtue of (3.37) and (3.38), we get

$$\mathcal{Q}(X^{h+1}-W^{h+1}) = u_h\mathcal{M}_{X,h}(X^h-X^{h+1}),\ b-\mathcal{A}(W^{h+1}) = u_h\mathbf{M}_{y,h}(y^h-y^{h+1}).$$

In what follows, we prove that $\{(X^h;y^h)\}_{h=1}^{\infty}$ is bounded.

We define the weighted distance between $(X^{h+1}; y^{h+1})$ and $(\widetilde{X}; \tilde{y})$ as

$$d_w^{h+1} = u_h(\|\widetilde{X} - X^{h+1}\|^2_{\mathcal{M}_{X,h}} + \|\tilde{y} - y^{h+1}\|^2_{\mathbf{M}_{y,h}}).$$

By step 5, for the descent steps, it follows $u_{h-1} \geq u_h \geq u_{min}$ for any h, combining with (1.3), we have

$$\begin{aligned} d_w^{h+1} \leq & u_h(\|\widetilde{X} - X^h\|^2_{\mathcal{M}_{X,h}} + 2(\langle \mathcal{Q}(X^{h+1} - W^{h+1}), \widetilde{X} - X^{h+1}\rangle + \langle \tilde{y} - y^{h+1}, b - \mathcal{A}(W^{h+1})\rangle) \\ \leq & d_w^h + 2(\langle \mathcal{Q}(X^{h+1} - W^{h+1}), \widetilde{X} - X^{h+1}\rangle + \langle \tilde{y} - y^{h+1}, b - \mathcal{A}(W^{h+1})\rangle). \end{aligned}$$

Moreover, it holds

$$\begin{aligned} &\langle \mathcal{Q}(X^{h+1} - W^{h+1}), \widetilde{X} - X^{h+1}\rangle - \langle \mathcal{Q}(X^{h+1} - \widetilde{X}), W^{h+1} - \widetilde{X}\rangle + \\ &\frac{1}{2}\langle X^{h+1} - \widetilde{X}, \mathcal{Q}(X^{h+1} - \widetilde{X})\rangle = -\frac{1}{2}\langle X^{h+1} - \widetilde{X}, \mathcal{Q}(X^{h+1} - \widetilde{X})\rangle \leq 0. \end{aligned}$$

Therefore, we obtain $d_w^{h+1} \leq d_w^h + 2(f_l(X^h; y^h) - f_{W^{h+1}}(X^{h+1}; y^{h+1}) + a\epsilon^h)$.

By virtue of (3.40), both $\{f_l(X^h; y^h)\}_{h=1}^{\infty}$ and $\{f_u(X^h; y^h)\}_{h=1}^{\infty}$ are nonincreasing.

Hence, for any fixed $H \geq 1$ and any $h > H$, we get

$$\text{(A.9)} \qquad d_w^{h+1} \leq d_w^H + 2\sum_{i=0}^{\infty}(f_l(X^i; y^i) + a\epsilon^i - f_{W^{i+1}}(X^{i+1}; y^{i+1})),$$

since $f_u(X^h; y^h) - f_{W^{h+1}}(X^{h+1}; y^{h+1}) \geq f_u(X^{h+1}; y^{h+1}) - f_{W^{h+1}}(X^{h+1}; y^{h+1}) \geq 0$.

Meanwhile,

$$\begin{aligned} &\sum_{h=0}^{\infty}(f_l(X^h; y^h) - f_l(X^{h+1}; y^{h+1}) - a\epsilon^{h+1}) \leq \sum_{h=0}^{\infty}(f(X^h; y^h) - f(X^{h+1}; y^{h+1})) \\ &= \lim_{n\to\infty}(f(X^0; y^0) - f(X^{n+1}; y^{n+1})) \leq f(X^0; y^0) - f(\widetilde{X}; \tilde{y}), \end{aligned}$$

then $\sum\limits_{h=0}^{\infty}(f_l(X^h; y^h) - f_l(X^{h+1}; y^{h+1}) - a\epsilon^{h+1}) =: M < \infty$.

In addition, using (3.39), it holds

$$\text{(A.10)} \qquad \begin{aligned} m_L\sum_{h=0}^{\infty}(f_l(X^h; y^h) - f_{W^{h+1}}(X^{h+1}; y^{h+1})) &\leq \sum_{h=0}^{\infty}(f_l(X^h; y^h) - f_u(X^{h+1}; y^{h+1})) \\ &\leq f(X^0; y^0) - f(\widetilde{X}; \tilde{y}). \end{aligned}$$

In the following, we shall prove that

$$\text{(A.11)} \qquad m_L\sum_{h=0}^{\infty}(f_l(X^h; y^h) + a\epsilon^h - f_{W^{h+1}}(X^{h+1}; y^{h+1})) =: M_1 < \infty.$$

According to (A.10), it suffices to prove that $a\sum_{h=0}^{\infty}\epsilon^h$ converges.

Note that

$$M = \sum_{h=0}^{\infty}(f_l(X^h; y^h) - f_u(X^{h+1}; y^{h+1})) = \lim_{n\to\infty}(f_l(X^0; y^0) - (f_l(X^{n+1}; y^{n+1}) + a\sum_{h=0}^{n}\epsilon^{h+1})),$$

thus $\lim_{n\to\infty}((f_l(X^{n+1}; y^{n+1}) + a\sum_{h=0}^{n}\epsilon^{h+1})) = f_l(X^0; y^0) - M$.

By the assumption, $0 \leq a\epsilon^{h+1} \leq \delta^h$ and $\{\delta^h\}$ converges, we conclude that $\{a\epsilon^h\}$ is bounded, say $0 \leq a\epsilon^h \leq c < \infty$ for all h.

Hence, together with (3.40) and (3.41), we have $f_l(X^0;y^0) \geq f_l(X^1;y^1) \geq \cdots \geq f(\widetilde{X};\tilde{y}) - c$, whence $\lim_{h\to\infty} f_l(X^{h+1};y^{h+1})$ exists, so does $a\sum_{h=0}^{\infty}\epsilon^h$, which in turn yields $\lim_{h\to\infty}\epsilon^h = 0$.

Therefore (A.9) and (A.11) imply that, for any $h > H$, $d_w^{h+1} \leq d_w^H + \frac{2M_1}{m_L}$, whence both $\{X^h\}$ and $\{y^h\}$ are bounded, and then $\{(X^h;y^h)\}$ has an accumulation point, say $(X^\star;y^\star)$, i.e.,

$$\exists\ \{X^{h_j}\} \text{ and } \{y^{h_j}\}, \text{ s.t. } X^{h_j} \to X^\star, y^{h_j} \to y^\star \text{ and } f(X^{h_j};y^{h_j}) \to f(X^\star;y^\star) \text{ as } j \to \infty.$$

Since $f(X^0;y^0) \geq f(X^1;y^1) \geq \cdots \geq f(X^h;y^h) \geq \cdots \geq f(\widetilde{X};\tilde{y})$, we get

$$f(X^h;y^h) \downarrow \inf f(X^h;y^h) \text{ as } h \to \infty.$$

Thus $f(X^\star;y^\star) = \inf f(X^h;y^h)$, whence $f(X^h;y^h) \geq f(X^\star;y^\star)$ for any h.

By the same argument as the above, we have, for any $h > H$,

$$\begin{aligned}&u_h(\|X^\star - X^{h+1}\|^2_{\mathcal{M}_{X,h}} + \|y^\star - y^{h+1}\|^2_{\mathbf{M}_{y,h}})\\ \leq& u_{H-1}(\|X^\star - X^H\|^2_{\mathcal{M}_{X,H-1}} + \|y^\star - y^H\|^2_{\mathbf{M}_{y,H-1}}) + 2\sum_{i=H}^{\infty}(f_l(X^i;y^i) + a\epsilon^i - f^i_{W^{i+1}}(X^{i+1};y^{i+1})).\end{aligned}$$

Therefore, for any arbitrary $\varepsilon > 0$, one can choose large enough $\widehat{H}$, such that for any $h > \widehat{H}$, $u_h(\|X^\star - X^{h+1}\|^2_{\mathcal{M}_{X,h}} + \|y^\star - y^{h+1}\|^2_{\mathbf{M}_{y,h}}) < \varepsilon$, whence $\lim_{h\to\infty}(X^h;y^h) = (X^\star;y^\star)$. Consequently, $\mathcal{Q}(X^{h+1} - W^{h+1}) = u_h\mathcal{M}_{X,h}(X^h - X^{h+1}) \to 0$ and $b - \mathcal{A}(W^{k+1}) = u_h\mathbf{M}_{y,h}(y^h - y^{h+1}) \to 0$ as $h \to \infty$, and then

$$\lim_{h\to\infty} c_h u_h \sqrt{l_d^h} = 0. \tag{A.12}$$

On the other hand, (A.11) results in

$$f_l(X^h;y^h) + a\epsilon^h - f_{W^{h+1}}(X^{h+1};y^{h+1}) \to 0 \text{ as } h \to \infty. \tag{A.13}$$

Note that $0 \leq f(X^h;y^h) - f_{W^{h+1}}(X^{h+1};y^{h+1}) \leq f_l(X^h;y^h) + a\epsilon^h - f_{W^{h+1}}(X^{h+1};y^{h+1})$, hence $f(X^h;y^h) - f_{W^{h+1}}(X^{h+1};y^{h+1}) \to 0$ as $h \to \infty$, and then $f_{W^{h+1}}(X^{h+1};y^{h+1}) \to f(X^\star;y^\star)$ as $h \to \infty$. Now the proof of the statement that $(X^\star;y^\star)$ is a minimizer of $f(X;y)$ is similar to what we have done for proving the the end of Proposition 3.5.

We now show that $\lim_{h\to\infty}\nu^h = 0$. Since

$$f_u(X^h;y^h) - f_{W^{h+1}}(X^{h+1};y^{h+1}) \geq f_u(X^{h+1};y^{h+1}) - f_{W^{h+1}}(X^{h+1};y^{h+1}) = \eta^{h+1} \geq 0,$$

putting together with (A.13), implies $\lim_{h\to\infty}\eta^h = 0$ hence $\lim_{h\to\infty}\nu^h = 0$ by (A.12). □

(†) Division of Mathematical Sciences, School of Physical & Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore
E-mail address: LINH0016@e.ntu.edu.sg

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