An inexact spectral bundle method and error bounds for convex quadratic symmetric cone programming

Lin, Hui Ling

2012

Lin, Hui Ling. (2012). An inexact spectral bundle method and error bounds for convex quadratic symmetric cone programming. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/48908

https://doi.org/10.32657/10356/48908

NANYANG TECHNOLOGICAL UNIVERSITY

AN INEXACT SPECTRAL BUNDLE METHOD AND ERROR BOUNDS FOR CONVEX QUADRATIC SYMMETRIC CONE PROGRAMMING

LIN HUILING

School of Physical and Mathematical Sciences

2012

An Inexact Spectral Bundle Method & Error Bounds For CQSCP

LIN HUILING

2012

AN INEXACT SPECTRAL BUNDLE METHOD AND ERROR BOUNDS FOR CONVEX QUADRATIC SYMMETRIC CONE PROGRAMMING

LIN HUILING

School of Physical and Mathematical Sciences

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctoral of Philosophy

2012

Acknowledgements

Even though a doctoral dissertation is perceived as an individual achievement, it is impossible to complete it without the help of many peoples.

Above all, my sincere gratitude goes to my supervisor Chua Chek Beng. I enjoyed the meeting almost every week that I had with my supervisor over the past three years. I had the pleasure of benefitting from his valuable guidance and continuous encouragement throughout my research work, his very deep knowledge of related fields, and his amazing intuition.

I am grateful to my classmate Wan Min for his help to improve my MATLAB code for calculating the symmetrized Kronecker product of two matrices in numerical experiments.

I am thankful to my classmate Timothy Foo Chong Hsian and my friend Yu Feng Ling for offering their help in checking on the grammatical correctness of part of my thesis statement.

I would also like to extend my gratitude to the professors and staff of the Division of Mathematical Sciences of School of Physical and Mathematical Sciences in the Nanyang Technological University for the stimulating courses and seminars, for their guidance and support. I would like to thank the University for providing excellent working conditions and the Scholarship which allowed me to undertake this research and complete my thesis.

I am thankful to all my friends, in particular, my officemates, for being the surrogate family during the years I stayed in Singapore and for their continuous moral support. I will always be appreciative of their friendship.

Finally, I would like to manifest my deepest gratitude and appreciation to my parents, my little sisters and my little brother for their unconditional love and persistent support they have given me all these years. And, last but not least, I am eternally indebted to my husband Zhen Huan for his love, understanding, endless patience and encouragement to complete this work.

Contents

Summary

In this dissertation, we apply the main technical tools from the theory of Jordan algebras to explore a conic programming problem that minimizes a convex quadratic function over the intersection of a symmetric cone with an affine subspace. Our adaptation is called convex quadratic symmetric cone programming (CQSCP). This optimization problem subsumes linear programming, quadratic programming, second-order cone programming and semidefinite programming as special cases.

Under a mild assumption, we reformulate this problem as the problem of minimizing the maximum eigenvalue of an affine function. We present an inexact spectral bundle method for solving this eigenvalue minimization problem. This method is a first-order method, hence it requires much less computational cost in each iteration than second-order approaches such as interior-point methods. While sometimes providing solutions of low accuracy, such method can attack large scale problems. In each iteration of our method, we compute the largest eigenvalue inexactly, and solve a small convex quadratic symmetric cone program 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 Lipschitzian error bound under a Slater type condition for the reformulation of the problem under consideration.

Another purpose of this dissertation is to investigate a Lipschitzian error bound for the CQSCP problem. We start by considering a sequence of strictly feasible solutions within a wide neighborhood of the central trajectory for the monotone symmetric cone linear complementarity problem (SCLCP). Under assumptions of strict complementarity and Slater's condition, we derive a necessary condition of a Lipschitzian error bound for the monotone SCLCP in general Euclidean Jordan algebras, and provide four different characterizations of such error bound in the setting of Euclidean Jordan

algebras embedded in associative algebras. Under Slater's condition, the monotone SCLCP is at least as general as CQSCPs. As a consequence, the error bound results for the monotone SCLCP apply to CQSCP.

Finally, we describe an application of our proposed method to convex quadratic semidefinite programming problems. Numerical experiments with matrices of order up to 2000 are performed, and the computational results establish the effectiveness of this method.

Notation

Scalars

δ_{ij}	$\equiv \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise;} \end{cases}$ the Kronecker delta
$\bar{n}$	$\equiv \frac{n(n+1)}{2}$
$\mathbb{R}$	the real line
$\mathbb{R}_{++}$	the real positive numbers
$\lvert\tau\rvert$	the absolute value of a scalar τ
τ_+	$\equiv \max\{0, \tau\}$; the positive part of a scalar

Sets

$arg \min_S f$	the set of constrained minimizers of f on S
$B(x, \gamma)$	the open ball in a Euclidean space with center x and radius γ
$\lvert S\rvert$	the cardinality of a finite set S
$conv(S)$	the convex hull of a set S
$\overline{S}, int(S)$	the topological closure and interior of a set S, respectively
$relint(S)$	the relative interior of a set S
τS	the coset $\{\tau s \mid s \in S\}$ of a set S in a Euclidean space and a scalar τ

Spaces

$\mathbb{R}^m$	the m-dimensional real vector space
$\mathbb{R}^{m \times n}$	the space of $m \times n$ real matrices
$\mathbb{S}^n$	the space of real symmetric matrices of order n
$\mathbb{S}^n_+$	the cone of positive semidefinite matrices in $\mathbb{S}^n$
$\mathbb{S}^n_{++}$	the cone of positive definite matrices in $\mathbb{S}^n$

Matrices

$\mathbf{0}$	matrix of all zeros of appropriate size
I	identity matrix of appropriate order
A^{-1}	the inverse of a matrix A
A^T	the transpose of a matrix A
$A \cdot B$	the Hadamard product of two matrices A and B in $\mathbb{R}^{m\times n}$
$A \succeq (\succ) B$	the partial ordering: $A - B \in \mathbb{S}^n_+ (\mathbb{S}^n_{++})$
$\|\|A\|\|_2$	$\equiv \sqrt{\lambda_{max}(A^TA)}$; the operator norm of $A \in \mathbb{R}^{m\times n}$, the largest singular value of A
$\|\|A\|\|_F$	$\equiv \sqrt{tr(A^TA)}$; the Frobenius norm of $A \in \mathbb{R}^{m\times n}$
$\lambda_{max}(A)$	the largest eigenvalue of a matrix $A \in \mathbb{S}^n$
$tr(A)$	$\equiv \sum_{i=1}^n A_{ii}$; the trace of a matrix $A = (A_{ij})_{n\times n} \in \mathbb{S}^n$
$Diag(y)$	the diagonal matrix with diagonal elements equal to the components of the vector y

Vectors

$\mathbf{1}$	vector of all ones of appropriate size
$\mathbf{1}_i$	vector of appropriate size with 1 in the i-th coordinate and 0's elsewhere
$\langle x, y\rangle$	$\equiv x^Ty$; the standard inner product of vectors x and y in $\mathbb{R}^m$
$\|\|x\|\|_2$	$\equiv \sqrt{x^Tx}$; the l_2-norm of $x \in \mathbb{R}^m$
$\|\|x\|\|_{\mathbf{M}}$	$\equiv \sqrt{x^T\mathbf{M}x}$; the norm induced by the matrix $\mathbf{M} \in \mathbb{S}^m_{++}$

Algebra

$* : \mathbb{A} \to \mathbb{A}$	involution on an algebra $\mathbb{A}$
$\mathbf{0}$	zero element in an algebra or zero mapping on an algebra; its precise meaning depends on the context
$\mathbb{C}$	the field of complex numbers
$\mathbb{A}$	power-associative or associative algebra, unless otherwise specified
$\mathbb{V}$	Euclidean Jordan algebra, unless otherwise specified
$\mathbb{K}$	the cone of squares in $\mathbb{V}$
$\lambda_{min}(x)$	the smallest eigenvalue of an element x in a power-associative algebra
$\lambda_{max}(x)$	the largest eigenvalue of an element x in a power-associative algebra
$\lambda_i(x)$	(for $1 \le i \le r$) the i-th largest eigenvalue of an element x in

a power-associative algebra with rank r

$\lambda(x)$	ordered vector of eigenvalues of an element $x \in \mathbb{V}$ (see Section 2.3)
$\lambda(x, \mathbb{U})$	ordered vector of eigenvalues of an element x, considered in the subalgebra $\mathbb{U} \subseteq \mathbb{V}$
$rk(\mathbb{A})$	the rank of an algebra $\mathbb{A}$
$rk(x)$	the rank of an element $x \in \mathbb{V}$
$tr(x)$	the trace of an element x in a power-associative algebra (see Section 2.1)
e	unit element in an algebra
xy	the associative product of two elements x and y in an associative algebra
$x \bullet y$	$\equiv tr(xy^*)$; the trace inner product of two elements x and y in an associative algebra
$u \circ v$	the Jordan product of two elements u and v in a Jordan algebra
$\langle u, v \rangle$	$\equiv tr(u \circ v)$; the trace inner product of two elements u and v in $\mathbb{V}$
$\|\|u\|\|_F$	$\equiv \sqrt{\langle u, u \rangle}$; the Frobenius norm of $u \in \mathbb{V}$
$\mathcal{I} : \mathbb{A} \rightarrow \mathbb{A}$	identity map on an algebra $\mathbb{A}$
$\mathcal{Q} : \mathbb{V} \rightarrow \mathbb{V}$	a self-adjoint positive semidefinite (resp. definite) linear operator on $\mathbb{V}$, namely, $\langle \mathcal{Q}(x), y \rangle = \langle x, \mathcal{Q}(y) \rangle$ and $\langle \mathcal{Q}(x), x \rangle \geq 0$(resp. > 0) for all $x, y \in \mathbb{V}$
$\mathbf{Q}$	the matrix representation of $\mathcal{Q}$ with respect to a given canonical orthonormal basis of $\mathbb{V}$
$tr(\mathcal{Q})$	$\equiv tr(\mathbf{Q})$; the trace of $\mathcal{Q}$, it is the sum of all eigenvalues of $\mathbf{Q}$
$\|\|\mathcal{Q}\|\|_2$	$\equiv \\|\mathbf{Q}\\|_2$; the operator norm of $\mathcal{Q}$,
$\|\|\mathcal{Q}\|\|_F$	$\equiv \|\|\mathbf{Q}\|\|_F$; the Frobenius norm of $\mathcal{Q}$
$\|\|x\|\|_{\mathcal{M}}$	$\equiv \sqrt{\langle x, \mathcal{M}(x) \rangle}$; the norm induced by the self-adjoint positive definite linear operator $\mathcal{M}$
L_x	the Lyapunov transformation for an element x in a Jordan algebra
P_x	the quadratic representation of an element x in a Jordan algebra
$P_{x,y}$	the polarized quadratic operator of two elements x and y in a Jordan algebra
$\{x, y, z\}$	$\equiv P_{x+z}(y) - P_x(y) - P_z(y) = 2P_{x,z}(y)$; the triple product in a Jordan algebra

Mapping and Functions

$\mathcal{A}^*$ the adjoint operator of a linear mapping $\mathcal{A}$

$dist(x, S)$ $\equiv \begin{cases} \inf_{y \in S} \|x - y\| & \text{if } S \text{ is nonempty,} \\ +\infty & \text{otherwise,} \end{cases}$ where $\|\cdot\|$ denotes an appropriate norm;
the distance from a point x to the set S in a normed linear space

$f \diamond g$ composition of two functions f and g

$f^\sharp(y)$ the conjugate of a convex function $f(x)$

$O(g(x))$ the class of all functions $f(x)$ such that $||f(x)|| \leq \xi g(x)$ for some constant $\xi > 0$ and all $x \in \tilde{S}$, where $f : S \to \mathbb{V}$, $g : S \to \mathbb{R}_{++}$, S is an arbitrary set, $\tilde{S} \subseteq S$, and $\|\cdot\|$ denotes an appropriate norm

$\Theta(g(x))$ the class of all functions $F(x)$ such that $F(x) = O(g(x))$ and $F(x)^{-1} = O(\frac{1}{g(x)})$ for all $x \in \tilde{S} \subseteq S$, where $F : S \to int(\mathbb{K})$ and S is an arbitrary set

Functional analysis

JF the Jacobian of a mapping F

$\partial f(x)$ $\equiv \{g \in \mathbb{R}^m \mid f(y) \geq f(x) + \langle g, y - x\rangle \text{ for all } y \in \mathbb{R}^m\}$;
the subdifferentiable of f at $x \in \mathbb{R}^m$

$\partial_\varepsilon f(x)$ $\equiv \{g \in \mathbb{R}^m \mid f(y) \geq f(x) + \langle g, y - x\rangle - \epsilon \text{ for all } y \in \mathbb{R}^m\}$;
the ε-subdifferentiable of f at $x \in \mathbb{R}^m$

Chapter 1

Introduction

Convex quadratic symmetric cone programming includes the important class of convex quadratic semidefinite programming (CQSDP) problems. The motivation for the work in this thesis emanates from the question of solving large scale CQSDPs with reasonable storage requirement. Since interior-point methods are inappropriate for large scale CQSDPs due to their memory requirements, one goal of this thesis is to propose an inexact spectral bundle method for a convex quadratic problem involving symmetric cones, which comprise positive orthants, second-order cones and positive semidefinite cones, among others. As an arbitrary symmetric cone can be visualized as the cone of squares in an associated Euclidean Jordan algebra, our approach enables one to study major classes of optimization problems (e.g. the standard convex quadratic programming, second-order cone programming and CQSDP) with the help of a simple and unifying Jordan-algebraic technique.

1.1 Convex quadratic symmetric cone programming

Let $\mathbb{V}$ be a finite-dimensional Euclidean Jordan algebra (see Section 2.3 for its definition) of rank r, endowed with the trace inner product

$$\langle u, v\rangle = tr(u \circ v)$$

for all $u, v \in \mathbb{V}$. The convex quadratic symmetric cone program (CQSCP) is

$$\text{(P)} \qquad \begin{array}{rl} \min & \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle + \langle \hat{s}, x\rangle =: h(x) \\ \text{s.t.} & \mathcal{A}(x) = b \\ & x \in \mathbb{K}, \end{array}$$

where $\mathbb{K}$ is the cone of squares of $\mathbb{V}$, $\mathcal{Q}$ is a self-adjoint positive semidefinite linear operator acting on $\mathbb{V}$, $\mathcal{A}$ is a linear operator from $\mathbb{V}$ to $\mathbb{R}^m$, $b \in \mathbb{R}^m$, $\hat{s} \in \mathbb{V}$, and the scalar product $\langle , \rangle$ is the aforementioned trace inner product. Let $\mathcal{A}^*$ be the adjoint of $\mathcal{A}$. The Lagrange dual is equivalent to the following problem, called the dual convex quadratic symmetric cone program.

$$\text{(D)} \qquad \begin{array}{rl} \min & \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle + b^T y =: g(x, y) \\ \text{s.t.} & z = \mathcal{Q}(x) + \mathcal{A}^*(y) + \hat{s} \\ & x \in \mathbb{V},\ y \in \mathbb{R}^m,\ z \in \mathbb{K}. \end{array}$$

Under a mild assumption

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

in Chapter 3, we show that the Lagrange dual (D) is equivalent to the following eigenvalue minimization problem with an appropriate choice of α.

$$\text{(Eigform)} \qquad \min_{x,y}\ \alpha\lambda_1(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)) + b^T y + \tfrac{1}{2}\langle x, \mathcal{Q}(x)\rangle =: f(x, y),$$

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

Primal-dual path-following interior point algorithms have been provided in the literature for solving the CQSCP (P), see, for instance, [24, 60]. The former discuses a short-step path-following algorithm for the CQSCP (P), while the later proposes an inexact primal-dual infeasible path-following algorithm for it. Both of them present the polynomial complexity bound in terms of the rank of the underlying Euclidean Jordan algebra. The paper [24] showed that the CQSCP (P) can be reformulated as a monotone symmetric cone linear complementarity problem (SCLCP), and vice versa. Therefore, methods for solving the monotone SCLCP can be used to solve the CQSCP (P), see, [75, Chapter 7] and [51, 52, 59, 105]. The main event in each iteration of interior point algorithms is to solve for a search direction from a linear system of equations. Since the direct monotone SCLCP reformulation will double the dimension of the linear system, interior point algorithms may be inefficient for large scale CQSCPs.

When $\mathcal{Q} = \mathbf{0}$, we arrive at the class of symmetric cone programming. This is considered in, for instance, [4, 18, 89, 90]. In particular, it includes linear semidefinite programming (SDP) problems if $\mathbb{K}$ is the cone of symmetric positive semidefinite matrices. The SDPs have been exploited to develop approximation algorithms for NP-hard combinatorial optimization problems, such as the max cut problem and the traveling salesman problem.

By restricting $\mathbb{V} = \mathbb{S}^n$ and $\mathbb{K} = \mathbb{S}^n_+$ in the general formulation of CQSCP in (P) and (D), we obtain a pair of primal-dual CQSDPs. Henceforth, by abuse of notation, we still denote the primal CQSDP by CQSDP (P). Note that the CQSDP (P) is a special case of convex quadratically constrained quadratic semidefinite programs (CQCQSDPs) proposed by Sun and Zhang in [97], where the constraints are quadratic and convex.

The CQSDP (P) has many practical applications in economics and engineering. It captures several well-studied problems in the literature as special cases. An example is the nearest correlation matrix problem, where given a data matrix $B \in \mathbb{S}^n$ and a self-adjoint operator $\mathcal{L}$ on $\mathbb{S}^n$, one wants to solve

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

It arises in finance [39] and machine learning [83]. Another example is the Euclidean distance matrix completion problem (see, e.g., [7]). Various methods have been developed for solving the nearest correlation matrix problem. Very recent works include [27, 84] and the references therein. The former presents a second-order algorithm—an inexact smoothing Newton method—to solve it and demonstrates high efficiency of the proposed method with numerical experiments. Incidentally, second-order techniques may be applied to bundle methods, see, e.g., [79]. Such methods have high accuracy but high computational cost. The latter uses an augmented Lagrangian dual-based approach which is quadratically convergent. For more applications of the convex quadratic semidefinite programming, we refer the readers to [100] and to the references therein.

The CQSDP (P) can be solved by several existing methods. It can be reformulated as either a standard semidefinite-quadratic-linear program (see [102]) by introducing a few additional linear constraints and variables, or a semidefinite linear complementarity problem (see [53]) via its KKT system. However, their computational costs leave a lot to be desired for large scale problems. Therefore it is indispensable to design algorithms specifically for large scale convex quadratic symmetric cone programs, which are amenable to take advantage of the specific structure of the problem. To the best of our knowledge, there is so far just a few such methods even for the special case CQSDP (P). A theoretical primal–dual potential reduction algorithm using NT directions was proposed in [78]. The authors suggested the use of the conjugate gradient method to compute an approximate search direction. The paper [100] proposed an inexact primal–dual path-following method with three classes of pre-conditioners for the augmented equation via the preconditioned symmetric quasi-minimal residual (PSQMR) iterative solver. It converges quickly under suitable nondegeneracy assumptions. In our numerical experiments, 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 [26]. In her thesis [107], Zhao designed a semismooth Newton-CG augmented Lagrangian method for large scale convex quadratic semidefinite programs. Note that all these methods are not first-order methods.

More recently, a modified alternating direction method for the more general case of CQCQSDPs was proposed in [97]. It is a first-order method. Its main idea is to reformulate CQCQSDP 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. The computation time for our numerical experiments will be comparable with those in [97].

1.2 Motivation and organization of the thesis

It is known that symmetric cones are completely classified as direct sums of the following five concrete types of simple symmetric cones (see, e.g., [23, Chapter V]): (1) the second-order cone, (2) the cone of positive-definite real symmetric matrices, (3) the cone of positive-definite complex Hermitian matrices, (4) the cone of positive-definite quaternion Hermitian matrices, and (5) the exceptional 27-dimensional octonion cone. By convex quadratic symmetric cone programming, we can deal with these conic quadratic programming problems in a unified way. Convex optimization problem that minimizes a convex quadratic function over the intersection of an affine subspace with the nonnegative orthant, the second-order cone or the cone of positive semidefinite matrices, is well-studied, see, e.g., [2, 10, 81, 100]. Complex semidefinite programming (CSDP) [104] is to minimize a linear function over a feasible set given as the intersection of an affine subspace and a cone of complex Hermitian positive semidefinite matrices. CSDP has been studied in [104, 106] to yield a good approximation ratio for a classical combinatorial optimization problem, known as the max-3-cut problem. Subsequently, the authors of [56] applied the rounding procedure used by [104] for CSDP to solve the problem of arranging elements on a circle so as to approximately preserve specified pairwise distances, which is closely related to optimization problems found in genome assembly. In addition, approximation algorithms were designed in [62] to obtain approximate ratio for the max 3-section problem via CSDP relaxation.

The code **SBmethod** is a C++ implementation of the spectral bundle method [36, 38] of nonsmooth optimization, which is publically available [34]. Based on the results of evaluating all computer codes submitted to the Seventh DIMACS Implementation Challenge on Semidefinite and Related Optimization Problems [71], **SBmethod** is very efficient for a special class of large scale linear SDP problems, such as max cut problems, min k-uncut problems from frequency assignment, min bisection problems from circuit partitioning and problems involving computing the theta function of Hamming graphs.

As mentioned in [97], first-order methods usually require much less computation per iteration, and therefore might be suitable for relatively large problems. Meanwhile, this type of method is relatively easy to implement but at the cost of a poor convergence rate, which will be illustrated by the performance results (see Chapter 7) of our proposed Algorithm 4.4 on the nearest correlation matrix problem. The **SBmethod** is a first-order algorithm and hence provides a plan of attack for large scale linear SDP problems. It is this main strength of **SBmethod** that spurs us to extend the spectral bundle method for convex quadratic symmetric cone programming problems.

We do not review the state-of-the-art of bundle methods but are inclined to provide some elementary references here. The work [67] is an excellent primer and [11, 42] are complete treatment of the subject. Furthermore, the work [66] addresses 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 presented in, for instance, [5, 8, 46, 48, 50, 77, 92] and the references therein. We emphasize that a bundle method has been employed to solve a quite general problem—the equilibrium problem in [77], which covers a wide range of problems, 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 increasingly used in many practical applications, for instance, in economics, optimal control and engineering, see, for example, [12, 21, 87, 98] and the references therein.

The spectral bundle method [36, 38] is a specialization of the proximal bundle method of Kiwiel [49] to the largest eigenvalue optimization problem, which is one of three popular types of first-order bundle methods. We refer the readers to [98] for realizing a classification of bundle methods.

In this thesis, we propose an inexact spectral bundle method to solve the large scale CQSCP (P). Our proposed method has the advantage of being simple and cheap in computational costs. At each iteration one only needs to compute a maximum eigenvalue and a small-sized subproblem, which is a convex quadratic symmetric cone program. It is "small-sized" in the sense that the rank of the underlying Euclidean

Jordan algebra is less than 8. A major difference between our method applied to CQSDP (P) and the spectral bundle method for linear semidefinite programming [38] is that the maximum eigenvalue problem is no longer sparse in the case where both the linear map $\mathcal{A}$ and the matrix $\hat{s}$ are sparse. To avoid the high cost of computing the maximum eigenvalue, we use an existing program **eigifp** [45] to compute it inexactly in our numerical experiments. Meanwhile, the subproblems can be solved efficiently by an interior-point algorithm.

The essential structure of our proposed Algorithm 4.4 is that of Algorithm 4.1 in [38]. Our proposed Algorithm 4.4 generates a sequence of *trial points* $\{(x^k, y^k)\}_{k=1}^{\infty}$, which contains a subsequence of *stability centers* $\{(\hat{x}^k, \hat{y}^k)\}_{k=1}^{\infty}$, starting from (x^0, y^0) and defined by

$$(x^{k+1}, y^{k+1}) = arg \min_{x,y} \left\{ f_{\widehat{\mathcal{W}}^k}(x, y) + \frac{\nu_k}{2}(\|x - \hat{x}^k\|^2_{\mathcal{M}_{x,k}} + \|y - \hat{y}^k\|^2_{\mathbf{M}_{y,k}}) \right\}, \qquad (1.2)$$

where $f_{\widehat{\mathcal{W}}^k}(x, y)$ is an approximation of $f(x, y)$ (see Section 4.2.1 for its definition), the weight $\nu_k > 0$ controls the effect of the proximity to the previous iterate, $\mathcal{M}_{x,k} : \mathbb{V} \to \mathbb{V}$ is a given self-adjoint positive definite linear operator defining the norm $\| \cdot \|^2_{\mathcal{M}_{x,k}} = \langle \cdot, \mathcal{M}_{x,k}(\cdot) \rangle$, and $\mathbf{M}_{y,k}$ is a positive definite matrix. When (x^{k+1}, y^{k+1}) is a good trial point in the sense that it brings significant decrease in the value of f, we shall say that a *serious* step or *descent* step is made. Otherwise, a *null* step occurs. As has been pointed out in [69], the underlying structure of bundle methods is that of the approximate proximal point method in Algorithm 4.4.1 in [69]. Moreover, [88] demonstrates that bundle methods for computing zeros of general maximal monotone operators can be cast as a special case in a certain class of hybrid proximal point algorithms, and proves a linear rate of convergence for bundle methods. In brief, bundle methods in the literature combine the cutting plane method and Moreau-Yosida regularization technique, which guarantee both descent and stability properties. They are nonsmooth optimization methods. In addition, an entropy-like proximal point algorithm was designed for minimizing a convex function subject to the nonnegative orthant in [20]. This algorithm is extended to solve the problem of minimizing a closed proper convex function subject to symmetric cone constraints in [17].

In this thesis, our proposed method is a triple generalization of the spectral bundle method of [38]. Firstly, it extends to an inexact setting. It is "inexact" in the sense that it does not require to compute the exact eigenvalues. Instead, the approximated eigenvalues satisfy a sequence of prescribed tolerances. Our method can be viewed as a dual approach in the sense that we will solve the equivalent form of the dual problem under a reasonable assumption. We show that the sequence of *stability centers* $\{(\hat{x}^k, \hat{y}^k)\}_{k=1}^{\infty}$ generated by our proposed algorithm either is unbounded or converges to an exact solution for the reformulated problem, if the error in eigenvalue approximation approaches zero; otherwise it converges to an approximate solution. In addition, our proposed method produces a sequence of points as a byproduct, whose accumulation point is an exact or approximate solution for the CQSCP (P).

Secondly, we extend the domain of f to $\mathbb{V} \times \mathbb{R}^m$, while [38] considered the case when f is defined on $\mathbb{R}^m$. It is the variable in a Jordan algebra that necessitates the use of $(\mathcal{M}_{x,k}, \mathbf{M}_{y,k})$. Theoretically, we establish that our proposed Algorithm 4.4 converges globally with the hypothesis that

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

at a descent step, while

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

at a null step, and the assumption that $\{\mathcal{M}_{x,k}\}$ converges to a self-adjoint positive definite linear operator and $\{\mathbf{M}_{y,k}\}$ converges to a positive definite matrix, as the number of null steps or descent steps approaches infinity. However, in terms of computational cost, we expect that a good 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 properties (1.3) and (1.4). However, in general, it may be difficult to solve the subproblems efficiently with this choice, regardless of how robust the strategy is for modifying ν_k. The reason is that this choice results in computing the inverse of the operator $\mathcal{Q} + \nu_k \mathcal{I}$, whose computational cost may be the same as that of solving the minimum eigenvalue problem (Eigform) for large-scale problems. In Chapter 7, we provide an efficient choice of $(\mathcal{M}_{x,k}, \mathbf{M}_{y,k})$.

Thirdly, we provide an orthonormalization process to create a system of mutually orthogonal primitive idempotents from a finite set of primitive idempotents, which includes a new approximate subgradient. We exploit this new system of primitive idempotents to generate the underlying subalgebra of $\widehat{\mathcal{W}}^{k+1}$. Our orthonormalization process can ensure that the updated model $\widehat{\mathcal{W}}^{k+1}$ possesses a desirable property to guarantee convergence of our proposed method.

The study of error bounds has attracted a lot of attention and found many important applications during recent years, see [82] and [22, Chapter 6] for excellent surveys. In particular, error bounds can not only give us help in designing solution methods for solving optimization problems, for instance, in providing an effective and reasonable termination criteria, but can also be used to analyze the convergence rate. Moreover, it can be exploited in the sensitivity analysis of the problems when their data is subject to perturbation, see, e.g. [29, 94].

Another aim of the thesis is to explore a more desirable kind of error bounds, a Lipschitzian error bound, which is an upper bound estimation of the distance from a given point in a specified set to the solution set of the problem in terms of some residual functions. More precisely, let $\mathbb{E}$ be a normed linear space, let S and T be two subsets of $\mathbb{E}$, and let $r : S \cup T \to (0, \infty)$ be a problem-specific residual function with the property that $r(x) = 0$ if and only if $x \in S$. It is well known that a *Lipschitizan error bound* holds for the pair (S, T) if there exists a constant $\tau > 0$ such that

$$dist(x, S) \leq \tau r(x) \text{ for all } x \in T. \tag{1.5}$$

When $r(x) = \varphi(x) - \min_{x \in T} \varphi(x)$ and $S = arg \min_{x \in T} \varphi(x)$, where φ is a lower semi-continuous proper convex function, the inequality (1.5) defines a notion of *weak sharp minima*, which has been much discussed in the literature; see [13–16]. In our context, using the same residual function, namely, $r(x, y) = f(x, y) - \inf_{x,y} f(x, y)$ but a different set $T = \{(x^k, y^k)\}_{k=1}^N$ with prescribed properties, we obtain a Lipschitzian error bound for the eigenvalue minimization problem (Eigform). This error bound, in turn, establishes the reasonability of the stopping rule in our proposed method.

In the thesis, we also study some necessary and sufficient conditions of a Lipschitzian error bound for the CQSCP (P) under the Slater condition

(**A2**) the CQSCP (P) is strictly feasible.

This study is motivated from the recent work [19], where the residual function is specified as duality gaps of primal-dual solutions that lie within a certain neighborhood of the central path of a pair of primal-dual semidefinite programming problems. Under strict complementarity and Slater's condition, the paper [19] proved that the analyticity of extended Cholesky weighted centers provides a sufficient condition for such Lipschitzian error bound, and provided two necessary and sufficient conditions for this error bound. Such Lipschitzian error bound is also discussed in [64,93]. As we mentioned earlier, the CQSCP (P) can be reformulated as a monotone symmetric cone linear complementarity problem. This reformulation offers us a much more tractable framework for the study of error bounds. We establish our error bound results by exploring the limiting behavior of a sequence of strictly feasible solutions within a wide neighborhood of central trajectory for the monotone SCLCP reformulation.

The structure of the remaining chapters is given as follows.

- In Chapter 2, we present several definitions, examples and results from the theory of Euclidean Jordan algebras, as well as some decomposition structures in associative algebras, used in this thesis.

- In Chapter 3, we consider the conic programming model of minimizing a convex quadratic function over the intersection of a symmetric cone with an affine subspace. We discuss a relationship of this problem with a well-known eigenvalue optimization problem of minimizing the maximum eigenvalue function. In addition, we review the necessary concepts of subgradients and approximate subgradients that are key notions in nonsmooth analysis, and introduce a Ritz-type triplet to define an inexact computation of the largest eigenvalue.

- In Chapter 4, we first recall some theoretical issues related to general bundle methods. Our proposed Algorithm 4.4 is a specialization of the general scheme for bundle methods. Since the description of our algorithm is quite involved, we start with conceptual comments, passing to technical details gradually. In particular, we design an orthonormalization process to update the model $\widehat{\mathcal{W}}^k$. We provide a proof of the global convergence of this algorithm using techniques from the analysis of the general bundle method.

- In Chapter 5, we study the limiting behavior of a sequence of stability centers $\{(\hat{x}^k, \hat{y}^k)\}_{k=1}^{\infty}$ as $k \to \infty$ in the situation when the error in eigenvalue approximation tends to a positive constant. We deduce that each accumulation point of this sequence is an approximate solution of the minimum eigenvalue problem (Eigform) with the prescribed requirements for the approximate largest eigenvalue computation and the choice of the weight ν_k. Likewise, under a Slater type condition, we explore a Lipschitzian error bound for the eigenvalue minimization problem (Eigform) to measure how good each (x^k, y^k) is. We should mention that the Weyl's inequality in general Euclidean Jordan algebras plays a crucial role in the analysis of such Lipschitzian error bound. In matrix theory, the well-known Weyl's inequality establishes relations between the eigenvalues of two Hermitian matrices with those of their sum. This result was generalized recently to the setting of simple Euclidean Jordan algebras in [72]. We further prove Weyl's inequality in general Euclidean Jordan algebras.

- In Chapter 6, we first discuss some properties of the solution set of a monotone symmetric cone linear complementarity problem, and then consider the limiting behavior of a sequence of strictly feasible solutions within a wide neighborhood of central trajectory for the monotone SCLCP. Under assumptions of strict complementarity and Slater's condition, we derive a necessary condition of a Lipschitzian error bound for the monotone SCLCP in general Euclidean Jordan algebras, and provide four different characterizations of such error bound in the setting of Euclidean Jordan algebras embedded in associative algebras. Thanks to the observation that the a pair of primal-dual CQSCPs is an instance of the monotone SCLCP, we obtain the same error bound results for the CQSCP (P) as a byproduct.

- In Chapter 7, to validate our approach, we describe a MATLAB implementation of the inexact spectra bundle method for computing large scale CQSDPs, and analyze its performance on several data sets of test problems.

- In Chapter 8, we sum up the major results of this thesis and discuss a few possible future work directions.

Chapter 2

Jordan Algebras

Jordan algebras sprung from the search for elucidating a particular aspect of physics, quantum-mechanical observables. Surprisingly, they came out to have illuminating connections with a very large spectrum of mathematics. The book [68] provides a colloquial survey of Jordan algebras.

In this chapter, we briefly present some basic concepts and the main technical tools that are of a constant use in the study of Jordan algebras. They include the characteristic polynomial, the Peirce decomposition theorems and the spectral decomposition theorems. We also need the Schur complement and associative Peirce decomposition to explore the error bound analysis. The readers can find a more detailed discussion on Jordan algebra theory, for instance, in [23] and [68].

2.1 Power-associative and associative algebras

An n-dimensional vector space $\mathbb{A}$ over the field $\mathbb{R}$ or $\mathbb{C}$ is called an algebra over $\mathbb{R}$ or $\mathbb{C}$ if it is equipped with a bilinear mapping $(x, y) \to xy$ (multiplication). If for any $x \in \mathbb{A}$ and any positive integers p and q, $x^p x^q = x^{p+q}$, then $\mathbb{A}$ is said to be *power associative*. Here x^p is defined recursively by $x^{p+1} = xx^p$, with $x^1 \equiv x$.

If for some element $e \in \mathbb{A}$, $xe = ex = x$ for every $x \in \mathbb{A}$, then e is called a unit element of $\mathbb{A}$. The unit element, if it exists, is unique. An algebra $\mathbb{A}$ does not necessarily have a unit element.

Let $\mathbb{A}$ be a power-associative algebra with the unit element e. For each $x \in \mathbb{A}$, let

$$deg(x) = \min\{k > 0 \mid \{e, x^1, \cdots, x^k\} \text{ are linearly dependent}\}.$$

Then $deg(x)$ is called the *degree* of the *minimal polynomial* of x, which is the monic polynomial p over the field $\mathbb{R}$ or $\mathbb{C}$ of least degree such that $p(x) = \mathbf{0}$. The *rank* of $\mathbb{A}$, $rk(\mathbb{A})$, is defined by

$$rk(\mathbb{A}) = \max\{deg(x) \mid x \in \mathbb{A}\}.$$

An element x is called *regular* if $deg(x) = rk(\mathbb{A})$.

Proposition 2.1. *[23, Proposition II.2.1] Let $\mathbb{A}$ be a power associative algebra with its canonical topology as a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, unit element e and rank r. The set of regular elements is open and dense in $\mathbb{A}$. There exist polynomials $a_1, a_2, \cdots, a_r$ on $\mathbb{A}$ such that the minimal polynomial of every regular element x is given by*

$$p(\lambda; x) = \lambda^r - a_1(x)\lambda^{r-1} + a_2(x)\lambda^{r-2} + \cdots + (-1)^r a_r(x).$$

The polynomials $a_1, a_2, \cdots, a_r$ are unique and a_j is homogeneous of degree j.

We define the characteristic polynomial of a regular element x to be its minimal polynomial, then we can continuously extend characteristic polynomials to all elements $x \in \mathbb{A}$ since the set of regular elements is dense. Hence, the characteristic polynomial $p(\lambda; x)$ is a degree r polynomial in λ. Its roots $\lambda_1, \lambda_2, \cdots, \lambda_r$ are the *eigenvalues* of x. The characteristic and minimal polynomials have the same set of roots, and the minimal polynomial has only simple roots. Thus the degree $deg(x)$ is the number of distinct eigenvalues of x.

We call

$$tr(x) := a_1(x) = \lambda_1 + \lambda_2 + \cdots + \lambda_r$$

and

$$det(x) := a_r(x) = \lambda_1 \lambda_2 \cdots \lambda_r$$

the *trace* and the *determinant* of x, respectively.

An involution, $*$, on an algebra $\mathbb{A}$ is a map which satisfies the following properties:

(1) $* : \mathbb{A} \to \mathbb{A}$ is linear and $(x^*)^* = x$,

(2) $(xy)^* = y^*x^*$,

(3) $tr(xx^*) \geq 0$.

If $x^* = x$, then x is self-adjoint. We denote $\mathbb{V} = \{x \in \mathbb{A} \mid x^* = x\}$, the set of self-adjoint elements of $\mathbb{A}$. Plainly, $\mathbb{V}$ is a subspace of $\mathbb{A}$.

Let $\mathbb{A}$ be a power-associative algebra with unit element e and involution $*$. We define the trace inner product as

$$x \bullet y = tr(xy^*) \text{ for all } \; x, y \in \mathbb{A}.$$

This inner product induces a norm given by $\|x\|_{F/\mathbb{A}} = \sqrt{tr(xx^*)}$.

An algebra $\mathbb{A}$ is associative if

$$(xy)z = x(yz) \; \text{ for all } x, y, z \in \mathbb{A}.$$

We further assume that $\mathbb{A}$ has the unit element e. An element x is said to be invertible if there is a unique element x^{-1} such that $xx^{-1} = x^{-1}x = e$. In an associative algebra with the unit element, it is easy to verify that for each x and invertible p, both x^* and pxp^{-1} have the same characteristic polynomial as x, and thus the same eigenvalues, trace, and determinant. Likewise, if x is invertible, then $xy = x(yx)x^{-1}$, thus xy and yx have the same characteristic polynomial.

2.2 Jordan algebras

Definition 2.1. *An algebra* $\mathbb{V}$ *over the field* $\mathbb{R}$ *or* $\mathbb{C}$ *is said to be a* Jordan algebra *if*

(i) $xy = yx$ *for all* $x, y \in \mathbb{V}$, *and*

(ii) $x(x^2y) = x^2(xy)$ *for all* $x, y \in \mathbb{V}$.

We also use $x \circ y$ to denote the product xy in a Jordan algebra, and call it the Jordan product. Jordan algebras are not necessarily associative, but they are power associative, see Proposition II.1.2 in [23]. Therefore, the definitions of rank, characteristic polynomial, eigenvalues, trace, and determinant remain valid for them.

Let $\mathbb{A}$ be an associative algebra with unit e and involution $*$. It is well known that the algebra $\mathbb{A}$ equipped with the product

$$x \circ y = \frac{xy + yx}{2}$$

is a Jordan algebra. This Jordan algebra is called the Jordan algebra *derived* from $\mathbb{A}$.

We call a Jordan algebra *special* if it is isomorphic to the Jordan algebra derived from an associative algebra; otherwise it is called *exceptional*.

Let $(\mathbb{V}, \circ)$ be a Jordan algebra. We assume throughout that there is a unit element $e \in \mathbb{V}$. We say that an element x is *invertible* if there exists a unique element y such that $x \circ y = e$. The inverse of x is denoted by x^{-1}. For each $u \in \mathbb{V}$, we write L_u for the *Lyapunov transformation* $L_u(v) := u \circ v$ for all $v \in \mathbb{V}$. Two elements $u, v \in \mathbb{V}$ are said to *operator commute* if $L_u L_v = L_u L_v$. The *quadratic representation* of u is $P_u := 2L_u^2 - L_{u^2}$. We define $P_{u,v} = L_u L_v + L_v L_u - L_{u \circ v}$. Note that L_u, P_u and $P_{u,v}$ are linear endomorphisms of $\mathbb{V}$. For each linear endomorphism A of $\mathbb{V}$, the trace of A is well defined, and denoted as $tr(A)$. We follow the definition from [68]: for $x, y, z \in \mathbb{V}$, the *triple product* $\{x, y, z\}$ is given by $P_{x+z}(y) - P_x(y) - P_z(y)$. Note that $\{x, y, z\} = 2P_{x,z}(y)$.

When $\mathbb{V}$ is the Jordan algebra derived from an associative algebra, we obtain for any $x, y, z \in \mathbb{V}$,

$$P_x(y) = xyx$$

and

$$P_{x,z}(y) = \frac{xyz + zyx}{2}.$$

2.3 Euclidean Jordan algebras

A Jordan algebra $(\mathbb{V}, \circ)$ is *Euclidean* if there exists a positive definite symmetric bilinear form $\langle \cdot, \cdot \rangle$ on $\mathbb{V}$ which is associative, that is

$$\langle x \circ y, z \rangle = \langle y, x \circ z \rangle \quad \text{for all } x, y, z \in \mathbb{V}.$$

This bilinear form is an inner product on $\mathbb{V}$. Since the bilinear form $tr(u \circ v)$ is symmetric positive definite and associative by [23, Proposition II 4.3], in the rest of this thesis we will assume that the inner product on $\mathbb{V}$ is given by the trace inner product

$$\langle u, v\rangle = tr(u \circ v)$$

for all $u, v \in \mathbb{V}$. Such algebras extend many properties of symmetric matrices, for instance, having real eigenvalues, the orthogonal eigenvectors, and the notion of Schur complement.

Note that L_u, whence P_u, is self-adjoint under $\langle \cdot, \cdot\rangle$ by [23, Proposition II 4.3].

Henceforth, to ease the writing, $\mathbb{V}$ shall denote a Euclidean Jordan algebra of finite dimension with unit e. Given a Euclidean Jordan algebra $\mathbb{V}$, its *cone of squares* is

$$\mathbb{K} = \{x^2 \mid x \in \mathbb{V}\}.$$

For an element $v \in \mathbb{V}$, we write $v \succeq \mathbf{0}$ if and only if $v \in \mathbb{K}$. By Theorem III 2.1 in [23], the interior $int(\mathbb{K})$ of $\mathbb{K}$ is a *symmetric cone*, i.e., $int(\mathbb{K})$ is a convex, homogeneous and self-adjoint cone. Thus $x \in \mathbb{K}$(resp., $x \in int(\mathbb{K})$) if and only if $\langle x, y\rangle \geq 0$(resp., $\langle x, y\rangle > 0$) for all $y \in \mathbb{K}$(resp., $\mathbf{0} \neq y \in \mathbb{K}$).

Example 2.1. *Let $\mathbb{S}^n$ denote the set of all $n \times n$ real symmetric matrices with the Jordan product defined by*

$$X \circ Y := \frac{1}{2}(XY + YX).$$

In this setting, $(\mathbb{S}^n, \circ)$ is a Euclidean Jordan algebra, and the cone of squares $\mathbb{S}^n_+$ is the set of all positive semidefinite matrices in $\mathbb{S}^n$. □

Example 2.2. *Let $\mathbb{L}^n$ be the n-dimensional real vector space whose elements x are indexed from zero, i.e.*

$$x = \begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix}.$$

The Jordan product is given by

$$x \circ y = \begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix} \circ \begin{bmatrix} y_0 \\ \bar{y} \end{bmatrix} := \begin{bmatrix} x^T y \\ x_0\bar{y} + y_0\bar{x} \end{bmatrix}.$$

The algebra $(\mathbb{L}^n, \circ)$ is a Euclidean Jordan algebra, a.k.a. the Jordan spin algebra. By the definition of the Jordan product, we have

$$x^2 = \begin{bmatrix} x_0^2 + \|\bar{x}\|_2^2 \\ 2x_0\bar{x} \end{bmatrix}.$$

Obviously, $x_0^2 + \|\bar{x}\|_2^2 \geq 2x_0\|\bar{x}\|$. Therefore, in this algebra, the cone of squares $\mathbb{L}^n_+$, called the Lorentz cone *(or the* second-order cone*), is*

$$\mathbb{L}^n_+ = \left\{ x \in \mathbb{R}^n \;\middle|\; \|\bar{x}\|_2 \leq x_0 \right\}.$$

□

We are now ready to list some fundamental properties of the operator P_u.

Proposition 2.2. *For elements $u, v, u_1, \cdots, u_k \in \mathbb{V}$, and an integer k, it holds*

(i) $P_u(e) = u^2$;

(ii) $P_{P_u(v)} = P_u P_v P_u$;

(iii) $P_{u^k} = (P_u)^k$;

(iv) if $u \in int(\mathbb{K})$, then $P_u(v) \in int(\mathbb{K})$;

(v) $P_{u_1+\cdots+u_k} = \sum_{i=1}^k P_{u_i} + 2\sum_{1\leq i<j\leq k} P_{u_i,u_j}$; and

(vi) the element u is invertible if and only if P_u is invertible, and $P_u^{-1} = P_{u^{-1}}$.

Proof. For (i), it is straightforward. For (ii), see [23], the proof of (iii) in Proposition II.3.3. For (iii), see [90], the proof of 5 in Lemma 8. For (iv), see [23], Proposition III.2.2 for a proof. For (v), the proof of the formula is a straightforward inductive proof . For (vi), see [23], the proof of Proposition II.3.1. □

Definition 2.2. *An idempotent of $\mathbb{V}$ is a nonzero element $c \in \mathbb{V}$ satisfying $c^2 = c$, it is a* primitive idempotent *if it cannot be written as a sum of two idempotents. Two idempotents c and d are said to be* orthogonal *if $c \circ d = \mathbf{0}$. Any idempotent is a sum of orthogonal primitive idempotents.*

We now review the *Peirce decomposition* with respect to an idempotent in a Euclidean Jordan algebra $\mathbb{V}$. Given an idempotent $c \in \mathbb{V}$, by virtue of Proposition III.1.3

of [23], the only possible eigenvalues of L_c are $1, \frac{1}{2}, 0$, therefore $\mathbb{V}$ is the direct sum of the corresponding subspaces $\mathbb{V}(c,1)$, $\mathbb{V}(c,\frac{1}{2})$ and $\mathbb{V}(c,0)$, where

$$\mathbb{V}(c,i) = \{x \in \mathbb{V} \mid L_c(x) = ix\}, \quad i \in \left\{1, \frac{1}{2}, 0\right\}.$$

This decomposition is called the *Peirce decomposition* of $\mathbb{V}$ with respect to the idempotent c. For any $x \in \mathbb{V}$, we write

$$x = x_1 + x_{\frac{1}{2}} + x_0$$

its Peirce decomposition with respect to c. We summarize some properties of the above three subspaces.

In what follows, for $\mathcal{A}, \mathcal{B}, \mathcal{C} \subseteq \mathbb{V}$, the notation $\mathcal{A} \circ \mathcal{B}$ means the set $\{x \circ y \mid x \in \mathcal{A}, y \in \mathcal{B}\}$, and $\{\mathcal{A}, \mathcal{B}, \mathcal{C}\}$ means the set $\{\{x,y,z\} \mid x \in \mathcal{A}, y \in \mathcal{B}, z \in \mathcal{C}\}$.

The Peirce spaces $\mathbb{V}(c,i)$ satisfy certain Peirce multiplication rules, which are stated in the following theorem.

Theorem 2.1. *Let c be an idempotent in a Euclidean Jordan algebra $\mathbb{V}$, we have the following rules for products of Peirce spaces $\mathbb{V}(c,i)$. We use the convention that $\mathbb{V}(c,l) = \{\mathbf{0}\}$ if $l \notin \{0, \frac{1}{2}, 1\}$.*

(i) $\mathbb{V}(c,1)$ *and* $\mathbb{V}(c,0)$ *are subalgebras of* $\mathbb{V}$ *and* $\mathbb{V}(c,1) \circ \mathbb{V}(c,0) = \{\mathbf{0}\}$.

(ii) $\mathbb{V}(c,\frac{1}{2}) \circ (\mathbb{V}(c,1) \oplus \mathbb{V}(c,0)) \subseteq \mathbb{V}(c,\frac{1}{2})$.

(iii) $\mathbb{V}(c,\frac{1}{2}) \circ \mathbb{V}(c,\frac{1}{2}) \subseteq \mathbb{V}(c,1) \oplus \mathbb{V}(c,0)$.

(iv) $\{P_u(v) \mid u \in \mathbb{V}(c,i), v \in \mathbb{V}(c,j)\} \subseteq \mathbb{V}(c, 2i-j)$ *for* $i,j \in \{0, \frac{1}{2}, 1\}$.

(v) $\{\mathbb{V}(c,i), \mathbb{V}(c,j), \mathbb{V}(c,k)\} \subseteq \mathbb{V}(c, i-j+k)$ *for* $i,j,k \in \{0, \frac{1}{2}, 1\}$.

Proof. For (i), (ii) and (iii), see the proof of Proposition IV 1.1 in [23]. The proof of (iv) can be found in both [68, Peirce Multiplication Theorem 8.2.1] and [31, Proposition 1]. For (v), see the proof in [68, Peirce Multiplication Theorem 8.2.1]. □

Let c be an idempotent in $\mathbb{V}$, we see that both $\mathbb{V}(c,1)$ and $\mathbb{V}(c,0)$ are Euclidean Jordan algebras by (i) of Theorem 2.1. We denote the cones of squares of $\mathbb{V}(c,1)$ and $\mathbb{V}(c,0)$ by $\mathbb{K}_1$ and $\mathbb{K}_0$, respectively.

For a square matrix given in the block form

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix},$$

where A_1 is square, if A_1 is invertible, the Schur complement of A_1 in A is defined by

$$A/A_1 = A_4 - A_3 A_1^{-1} A_2.$$

The paper [31] studied the concept of *Schur complement* in the setting of Euclidean Jordan algebras. The properties of Schur complement will serve as one of our basic ingredients of error bound analysis for convex quadratic symmetric cone programming. We follow the notation and state some elementary results of [31].

Definition 2.3. *Let c be an idempotent in $\mathbb{V}$, for $x \in \mathbb{V}$, we write $x = x_1 + x_{\frac{1}{2}} + x_0$ its Peirce decomposition with respect to c. When x_1 is invertible in $\mathbb{V}(c,1)$ with inverse x_1^{-1}, the* Schur complement *of x_1 in x is defined by*

$$x/x_1 := x_0 - P_{x_{\frac{1}{2}}}(x_1^{-1}).$$

Likewise, if x_0 is invertible in $\mathbb{V}(c,0)$ with inverse x_0^{-1}, the Schur complement *of x_0 in x is defined by*

$$x/x_0 := x_1 - P_{x_{\frac{1}{2}}}(x_0^{-1}).$$

Proposition 2.3. *Let c be an idempotent in $\mathbb{V}$, for $x \in \mathbb{V}$, we write $x = x_1 + x_{\frac{1}{2}} + x_0$ its Peirce decomposition with respect to c. The following hold.*

(i) If $x_1 \in int(\mathbb{K}_1)$, then $x \in int(\mathbb{K})$(resp., $x \in \mathbb{K}$) if and only if $x/x_1 \in int(\mathbb{K}_0)$ (resp., $x/x_1 \in \mathbb{K}_0$).

(ii) If $x_0 \in int(\mathbb{K}_0)$, then $x \in int(\mathbb{K})$(resp., $x \in \mathbb{K}$) if and only if $x/x_0 \in int(\mathbb{K}_1)$ (resp., $x/x_0 \in \mathbb{K}_1$).

Proof. See the proofs of [31, Proposition 1] and [31, Corollary 5]. □

Definition 2.4. *A* complete system *of $\mathbb{V}$ is a set of orthogonal idempotents $\{c_1, \cdots, c_k\}$ of $\mathbb{V}$ such that*

$$c_i \circ c_j = \mathbf{0} \quad \text{if} \;\; i \neq j, \quad \text{and} \quad c_1 + \cdots + c_k = e.$$

A Jordan frame *is a complete system of primitive idempotents.*

The number of elements in any Jordan frame is an invariant, which equals to the rank of $\mathbb{V}$ (see [23, p.45]). Note that $\langle c_i, c_j\rangle = \langle c_i \circ c_j, e\rangle = 0$ whenever $i \neq j$.

Henceforth, r shall denote the rank of $\mathbb{V}$.

We recall the following arguments concerning the operator commutativity.

Proposition 2.4. *[23, Lemma X.2.2] For $x, y \in \mathbb{V}$, they operator commute if and only if there exists a Jordan frame $\{c_1, \cdots, c_r\}$, such that $x = \sum_{i=1}^r \zeta_i c_i$ and $y = \sum_{i=1}^r \eta_i c_i$ with real numbers ζ_i, η_i.*

Proposition 2.5. *[32, Proposition 6] For $x, y \in \mathbb{V}$, the following conditions are equivalent:*

(i) $x, y \in \mathbb{K}$ and $\langle x, y\rangle = 0$.

(ii) $x, y \in \mathbb{K}$ and $x \circ y = \mathbf{0}$.

In each case, elements x and y operator commute.

Theorem 2.2. *(Spectral decomposition of Type I (Theorem III.1.1, [23])) For each $x \in \mathbb{V}$, there exist a unique complete system $\{c_1, \cdots, c_k\}$ and unique real numbers $\lambda_1 > \cdots > \lambda_k$ such that*

$$x = \sum_{i=1}^{k} \lambda_i c_i.$$

Theorem 2.3. *(Spectral decomposition of Type II (Theorem III.1.2, [23])) For each $x \in \mathbb{V}$, there exist a Jordan frame $\{c_1, \cdots, c_r\}$ and real numbers $\lambda_1(x) \geq \cdots \geq \lambda_r(x)$ such that*

$$x = \sum_{i=1}^{r} \lambda_i(x) c_i.$$

The numbers $\lambda_j(x)$ (with their multiplicities) are uniquely determined by x and they are the eigenvalues of x.

Henceforth, all spectral decompositions are of type II, unless otherwise stated.

By convention, for $x \in \mathbb{V}$, the coefficients $\lambda_1(x), \cdots, \lambda_r(x)$ in Theorem 2.3 are always arranged in decreasing order. In this sense, we define the eigenvalue map $\lambda : \mathbb{V} \to \mathbb{R}^r$ by

$$\lambda(x) = (\lambda_1(x), \cdots, \lambda_r(x))^T.$$

If x is invertible, the inverse is

$$x^{-1} = \lambda_1^{-1}(x)c_1 + \cdots + \lambda_r^{-1}(x)c_r.$$

In light of [95], each function $\lambda_i(x)$ is a Lipschitz continuous function. Note that $x \in \mathbb{K}$ (resp., $x \in int(\mathbb{K})$) if and only if $\lambda_i(x) \geq 0$ (resp., $\lambda_i(x) > 0$) for any $i = 1, \cdots, r$, see Theorems III.2.1 and III.3.1 of [23]. In this case, we define the square root of x as

$$x^{\frac{1}{2}} = \lambda_1^{\frac{1}{2}}(x)c_1 + \cdots + \lambda_r^{\frac{1}{2}}(x)c_r.$$

We adopt the notation provided in [31] and write $rk(x)$ for the *rank* of x, which is the number of nonzero eigenvalues of x. Then $rk(x) \leq rk(\mathbb{V})$ with equality holding if x is invertible. We denote by $\|x\|_{F/\mathbb{V}} := \sqrt{\sum_{i=1}^{r} \lambda_i^2(x)}$ the Frobenius norm induced by the trace inner product.

Corollary 2.1. *For an element $c \in \mathbb{V}$, the following statements are equivalent:*

(i) c is a primitive idempotent,

(ii) c is an idempotent and $dim(\mathbb{V}(c, 1)) = 1$,

(iii) $\lambda(c) = (1, 0, \cdots, 0)^T$.

Consequently, primitive idempotents have unit norm.

Proof. For the equivalence of (i) and (ii), see paragraph immediately after [23, Proposition IV. 1.2]. By Theorem 2.3, we see that (iii) implies (i). The second paragraph after [6, Theorem 5] states that (i) implies (iii). Thus, by the definition of trace, we have $\|c\|_F = \sqrt{tr(c \circ c)} = \sqrt{tr(c)} = \sqrt{\lambda_1(c) + \cdots + \lambda_r(c)} = 1$. □

Remark 2.1. *Let $\mathbb{A}$ be an associative algebra, if its Jordan subalgebra $\mathbb{V}$ of self-adjoint elements is Euclidean, then for any $x, y \in \mathbb{V}$, $x \bullet y = \langle x, y \rangle$. Moreover $||x||_{F/\mathbb{A}} = ||x||_{F/\mathbb{V}}$. We will drop subscripts with respect to algebras when the underlying algebra is clear from the context.*

The following proposition is a generalization of a trace inequality of *John von Neumann* in the framework of Euclidean Jordan algebras.

Proposition 2.6. *Let $u, v \in \mathbb{V}$. We have $tr(u \circ v) \leq \sum_{i=1}^{r} \lambda_i(u)\lambda_i(v) \leq \|u\|_F \|v\|_F$.*

Proof. For the first inequality, see [6], Theorem 23. The second inequality is straightforward. □

The next result tells us the sub-multiplicity of the Frobenius norm in associative algebras whose Jordan subalgebras of self-adjoint elements are Euclidean.

Lemma 2.1. *Let $\mathbb{A}$ be an associative algebra with an involution $*$ whose Jordan subalgebra $\mathbb{V}$ of self-adjoint elements is Euclidean. If $x, y \in \mathbb{A}$, then*

$$||xy||_F \leq ||x||_F ||y||_F.$$

Proof. Using Lemma 3 of [89], we have $xx^*, yy^* \in \mathbb{K}$. Let $yy^* = \sum_{i=1}^r \lambda_i c_i$ be a spectral decomposition of yy^*, then

$$\begin{aligned} ||xy||_F^2 = tr(xyy^*x^*) &= \sum_{i=1}^r \lambda_i tr(xc_ix^*) = \sum_{i=1}^r \lambda_i tr((xc_i)(xc_i)^*) \\ &\leq \lambda_1 \sum_{i=1}^r tr(xc_ix^*) = \lambda_1 tr(xx^*) \leq ||x||_F^2 ||y||_F^2, \end{aligned}$$

where the last inequality follows from the fact that $||y||_F^2 = \sum_{i=1}^r \lambda_i$. □

In this exposition, we also need to discuss the spectrum of elements x that lie in a Euclidean Jordan subalgebra $\mathbb{U}$ of $\mathbb{V}$. Note that the eigenvalue map of x depends on the algebra in which x is considered, we write this dependence explicitly as $\lambda(x, \mathbb{U})$ for its eigenvalue map in $\mathbb{U}$. Likewise, we denote by $\lambda_i(x, \mathbb{U})$ the ith eigenvalue of x in the subalgebra $\mathbb{U}$, $1 \leq i \leq rk(\mathbb{U})$.

Lemma 2.2. *[90, Lemma 13] Let $x \in \mathbb{V}$, then we obtain the smallest and the largest eigenvalue as*

$$\lambda_{min}(x) = \min_{\mathbf{0} \neq u \in \mathbb{V}} \frac{\langle u, x \circ u \rangle}{\langle u, u \rangle}, \ \lambda_{max}(x) = \max_{\mathbf{0} \neq u \in \mathbb{V}} \frac{\langle u, x \circ u \rangle}{\langle u, u \rangle}.$$

Proposition 2.7. *Let c be an idempotent of $\mathbb{V}$. The following statements hold:*

(i) If $u \in \mathbb{V}(c, \frac{1}{2})$, then $tr(u) = 0$.

(ii) If $\tilde{c}$ is a primitive idempotent in $\mathbb{V}(c, 1)$, then $\tilde{c}$ is a primitive idempotent of $\mathbb{V}$.

(iii) For $x \in int(\mathbb{K})$, we write $x = x_1 + x_{\frac{1}{2}} + x_0$ its Peirce decomposition with respect to c, then x_1 and x_0 have their inverses in $int(\mathbb{K}_1)$ and $int(\mathbb{K}_0)$, respectively.

(iv) Let $\tilde{r}$ be the rank of $\mathbb{V}(c,1)$ and $u \in \mathbb{V}(c,1)$, then

$$\{\lambda_1(u), \cdots, \lambda_r(u)\} = \{\lambda_1(u, \mathbb{V}(c,1)), \cdots, \lambda_{\tilde{r}}(u, \mathbb{V}(c,1)), 0, \cdots, 0\}$$

with $r - \tilde{r}$ zeros. If, in addition, $u \in \mathbb{K}$, then $\lambda_i(u) = \lambda_i(u, \mathbb{V}(c,1))$, $i = 1, \cdots, \tilde{r}$, and $\lambda_i(u) = 0$, $i = \tilde{r}+1, \cdots, r$.

Proof. Regarding (i): Since any idempotent is a sum of two orthogonal idempotents, there exists some idempotent d orthogonal to c such that $e = c+d$. Plainly, $u = 2c \circ u$, it yields $\langle u, d\rangle = 0$ and $tr(u) = \langle u, e\rangle = \langle u, c\rangle = 2\langle c \circ u, e\rangle = 2\langle u, c\rangle$, whence $tr(u) = 0$. For (ii), see the statement immediately after Proposition 2 of [25].

(iii) Since $\mathbb{V}(c,1)$ is a Euclidean Jordan algebra, by Lemma 2.2, we have

$$\begin{aligned}
\lambda_{min}(x_1, \mathbb{V}(c,1)) &= \min_{u \in \mathbb{V}(c,1)} \frac{\langle u, x_1 \circ u\rangle}{\langle u, u\rangle} \\
&= \min_{u \in \mathbb{V}, u_{\frac{1}{2}} = \mathbf{0}, u_0 = \mathbf{0}} \frac{\langle u, x \circ u\rangle}{\langle u, u\rangle} \geq \min_{u \in \mathbb{V}} \frac{\langle u, x \circ u\rangle}{\langle u, u\rangle} \\
&= \lambda_{min}(x) > 0,
\end{aligned}$$

where the last inequality follows from $x \in int(\mathbb{K})$. Similarly, we get $\lambda_{min}(x_0, \mathbb{V}(c,0)) > 0$.

Now to prove (iv). We use the trick mentioned in the statement after the proof of Corollary 34 in [6]. Suppose that $u = \sum_{i=1}^{r} \lambda_i(u) c_i$ is a spectral decomposition of u in $\mathbb{V}$, and $u = \sum_{i=1}^{\tilde{r}} \lambda_i(u, \mathbb{V}(c,1)) \tilde{c}_i$ is a spectral decomposition of u in the subalgebra $\mathbb{V}(c,1)$. By (ii), $\tilde{c}_1, \cdots, \tilde{c}_{\tilde{r}}$ are mutually orthogonal primitive idempotents in $\mathbb{V}$. It is possible to complete the system $\{\tilde{c}_1, \cdots, \tilde{c}_{\tilde{r}}\}$ to a complete system $\{\tilde{c}_1, \cdots, \tilde{c}_{\tilde{r}}, \tilde{c}_{\tilde{r}+1}, \cdots, \tilde{c}_r\}$ in $\mathbb{V}$. Note that we can arrange all the elements in $\{\lambda_1(u, \mathbb{V}(c,1)), \cdots, \lambda_{\tilde{r}}(u, \mathbb{V}(c,1)), 0, \cdots, 0\}$ in decreasing order, denoted by $\eta_1 \geq \cdots \geq \eta_r$, and rearrange $\{\tilde{c}_1, \cdots, \tilde{c}_{\tilde{r}}, \tilde{c}_{\tilde{r}+1}, \cdots, \tilde{c}_r\}$ in $\{\tilde{c}'_1, \cdots, \tilde{c}'_r\}$ such that

$$u = \sum_{i=1}^{\tilde{r}} \lambda_i(u, \mathbb{V}(c,1)) \tilde{c}_i = \sum_{i=1}^{r} \eta_i \tilde{c}'_i.$$

By virtue of Theorem 2.3, we have $\eta_i = \lambda_i(u),\ i = 1, \cdots, r$ and $\sum_{\{j|\eta_j=\lambda_j(u)\}} \tilde{c}'_j = \sum_{\{j|\eta_j=\lambda_j(u)\}} c_j$. If $u \in \mathbb{K}$, then $\eta_i = \lambda_i(u, \mathbb{V}(c, 1)),\ i = 1, \cdots, \tilde{r}$. The proof is complete. □

We now turn to the topic of the Peirce decomposition with respect to a Jordan frame.

Given a Jordan frame $\{c_1, \cdots, c_r\}$ in a Euclidean Jordan algebra $\mathbb{V}$, for $i, j \in \{1, \cdots, r\}$, we consider the following subspaces of $\mathbb{V}$

$$\begin{aligned} \mathbb{V}_{ii} &= \mathbb{V}(c_i, 1) = \{x \in \mathbb{V} \mid x \circ c_i = x\} = \mathbb{R}c_i, \\ \mathbb{V}_{ij} &= \mathbb{V}(c_i, \frac{1}{2}) \cap \mathbb{V}(c_j, \frac{1}{2}) = \left\{ x \in \mathbb{V} \;\middle|\; x \circ c_i = \frac{1}{2}x = x \circ c_j \right\}, \quad i \neq j. \end{aligned}$$

We have the following Peirce decomposition theorem with respect to a Jordan frame.

Theorem 2.4. *(Peirce Decomposition (Theorem IV.2.1, [23])) Given a Jordan frame $\{c_1, \cdots, c_r\}$, the space $\mathbb{V}$ decomposes in the following orthogonal direct sum:*

$$\mathbb{V} = \bigoplus_{1 \le i \le j \le r} \mathbb{V}_{ij}.$$

Furthermore,

$$\begin{aligned} &\mathbb{V}_{ij} \circ \mathbb{V}_{ij} \subseteq \mathbb{V}_{ii} + \mathbb{V}_{jj}, \\ &\mathbb{V}_{ij} \circ \mathbb{V}_{jk} \subseteq \mathbb{V}_{ik}, \quad \text{if } i \neq k, \\ &\mathbb{V}_{ij} \circ \mathbb{V}_{kl} = \{\mathbf{0}\}, \quad \text{if } \{i, j\} \cap \{k, l\} = \emptyset. \end{aligned}$$

Therefore we can write any element $x \in \mathbb{V}$ as

$$x = \sum_{i=1}^{r} x_i c_i + \sum_{1 \le i < j \le r} x_{ij},$$

with $x_i c_i = P_{c_i}(x)$ and $x_{ij} = 4L_{c_i}(L_{c_j}(x))$. This expression is called the Peirce decomposition of x with respect to the Jordan frame $\{c_1, \cdots, c_r\}$.

We note that in the above theorem, $x_i = \langle x, c_i \rangle,\ i = 1, \cdots, r$. Moreover, by (i) of Proposition 2.7, it holds $tr(x) = \sum_{i=1}^{r} x_i$.

In the following we fix a Jordan frame $\{c_1, \cdots, c_r\}$. Let P_{ij} be the orthogonal projection operator onto the subspace $\mathbb{V}_{ij}$. Then, by Theorem 2.4, we obtain

$$P_{ii} = P_{c_i}, \text{ and } P_{ij} = 4L_{c_i}L_{c_j} = 4L_{c_j}L_{c_i} = P_{ji}, \ i, j = 1, \cdots, r.$$

The orthogonal projection operators $\{P_{ij} \mid i, j = 1, \cdots, r\}$ satisfy

$$P_{ij}P_{kl} = \delta_{ik}\delta_{jl}P_{ij}, \quad \sum_{1 \leq i \leq j \leq r} P_{ij} = \mathcal{I}. \tag{2.1}$$

For details, see [68, Peirce Decomposition Theorem 13.14].

In what follows, let $x \in \mathbb{K}$ and $x = \sum_{i=1}^{r} \lambda_i(x)c_i$ be a spectral decomposition of x, we set

$$B := \{i \mid \lambda_i(x) > 0\}, \ N := \{i \mid \lambda_i(x) = 0\},$$

then $B \cup N = \{1, \cdots, r\}$. We define three projection operators as

$$P_{BB} = \sum_{i \leq j, i, j \in B} P_{ij}, \ P_{BN} = \sum_{i \in B, j \in N} P_{ij}, \ P_{NN} = \sum_{i \leq j, i, j \in N} P_{ij}.$$

In light of (2.1), the above three projection operators are orthogonal. Also, we specify the following three subspaces as

$$\mathbb{V}_{BB} = \sum_{i \leq j, i, j \in B} \mathbb{V}_{ij}, \ \mathbb{V}_{BN} = \sum_{i \in B, j \in N} \mathbb{V}_{ij}, \ \mathbb{V}_{NN} = \sum_{i \leq j, i, j \in N} \mathbb{V}_{ij}. \tag{2.2}$$

Since $\mathbb{V}_{ij} = P_{ij}\mathbb{V}$, we obtain

$$\mathbb{V}_{BB} = P_{BB}\mathbb{V}, \ \mathbb{V}_{BN} = P_{BN}\mathbb{V}, \ \mathbb{V}_{NN} = P_{NN}\mathbb{V}.$$

By virtue of Theorem 2.3 and Proposition 2.1 in [55], we have the following proposition.

Proposition 2.8. *Let $\mathbb{V}$ be a Euclidean Jordan algebra with rank r, then $\mathbb{V}$ is the orthogonal direct sum of three subspaces*

$$\mathbb{V} = \mathbb{V}_{BB} \oplus \mathbb{V}_{BN} \oplus \mathbb{V}_{NN}, \tag{2.3}$$

and for each $x \in \mathbb{V}$,

$$x = x_{BB} + x_{BN} + x_{NN} \tag{2.4}$$

where $x_{BB} \in \mathbb{V}_{BB}, x_{BN} \in \mathbb{V}_{BN}, x_{NN} \in \mathbb{V}_{NN}$, *and* $\mathbb{V}_{BB}, \mathbb{V}_{BN}, \mathbb{V}_{NN}$ *are specified by (2.2).*

For $1 \le k \le r$, we denote by

$$e_k = c_1 + \cdots + c_k,$$
$$\mathbb{V}^{(k)} = \mathbb{V}(e_k, 1) = \{x \in \mathbb{V} \mid x \circ e_k = x\}.$$

We recall some properties of this space in the following proposition, whose proof can be found in [33, Proposition 4.1].

Proposition 2.9. *The following statements hold:*

(i) $\mathbb{V}^{(k)}$ *is a subalgebra of* $\mathbb{V}$ *with rank* k, *its unit element is* e_k.

(ii) $\mathbb{V}^{(k)} = \mathbb{V}_{11} + \cdots + \mathbb{V}_{kk} + \sum_{1 \le i < j \le k} \mathbb{V}_{ij}$.

(iii) The closure of the symmetric cone of $\mathbb{V}^{(k)}$ *is equal to* $\mathbb{V}^{(k)} \cap \mathbb{K} =: \mathbb{K}^{(k)}$ *and is a face of* $\mathbb{K}$.

(iv) Let $P^{(k)}$ *denote the orthogonal projection from* $\mathbb{V}$ *onto* $\mathbb{V}^{(k)}$, *then* $P^{(k)} = P_{e_k}$.

For any finite set of mutually orthogonal primitive idempotents $\{a_1, \cdots, a_k\}$, $1 \le k \le r$, it is possible to complete this set to a Jordan frame. The above proposition yields that $\{x \in \mathbb{V} \mid x \circ (a_1 + \cdots + a_k) = x\}$, which by abuse of notation we still denote it by $\mathbb{V}^{(k)}$, is a subalgebra of $\mathbb{V}$ with rank k, its unit element is $a_1 + \cdots + a_k$, and the closure of its symmetric cone is $\mathbb{K} \cap \mathbb{V}^{(k)}$.

Proposition 2.10. *Let* $\mathbb{V}$ *be a Euclidean Jordan algebra with rank* r, $\mathbb{V}_{BB}, \mathbb{V}_{BN}, \mathbb{V}_{NN}$ *be specified by (2.2),* $|B| = k$ *and* $e_k = c_1 + \cdots + c_k$. *Then* $\mathbb{V}_{BB} = \mathbb{V}(e_k, 1)$ *and* $\mathbb{V}_{NN} = \mathbb{V}(e_k, 0)$ *are Euclidean Jordan algebras with rank* k *and* $r - k$ *and the closure of symmetric cones* $\mathbb{K}^{(k)}$ *and* $\mathbb{K}^{(r-k)}$, *respectively. Moreover,* $\mathbb{V}_{BN} = \mathbb{V}(e_k, \frac{1}{2})$.

Proof. It is straightforward from Propositions 2.9 and 2.8. □

A Jordan algebra is said to be *simple* if it does not contain any nontrivial ideal. Given a Euclidean Jordan algebra $\mathbb{V}$, we can write it as a direct sum of simple ideals (see [23, Proposition III.4.4]): $\mathbb{V} = \mathbb{V}_1 \oplus \cdots \oplus \mathbb{V}_\kappa$, each $\mathbb{V}_i$ is a Euclidean Jordan algebra. Thus any primitive idempotent in $\mathbb{V}$ is of the form $(\mathbf{0}, \cdots, \mathbf{0}, c^{(i)}, \mathbf{0}, \cdots, \mathbf{0})$ for some primitive idempotent $c^{(i)} \in \mathbb{V}_i$.

Example 2.3. *Consider $\mathbb{R}^n (n > 1)$, where any element x is written as $x = (x_1, \cdots, x_n)^T$. The inner product $\langle \cdot, \cdot \rangle$ is the usual dot product. The Jordan product is given by*

$$x \circ y = (x_1, \cdots, x_n)^T \circ (y_1, \cdots, y_n)^T := (x_1 y_1, \cdots, x_n y_n)^T.$$

It is easily verified that $(\mathbb{R}^n, \circ)$ is a Euclidean Jordan algebra. In this setting, all the non-trivial idempotents are $\mathbf{1}_i, i = 1, \cdots, n$. We see that $\mathbb{V}(\mathbf{1}_i, \frac{1}{2}) = \{\mathbf{0}\}, i = 1, \cdots, n$, whence it is not simple by Proposition IV 1.2 in [23]. □

We close this section with the classification theorem of Euclidean Jordan algebras.

Theorem 2.5. *Let $\mathbb{V}$ be a simple Euclidean Jordan algebra with rank $r \geq 2$, then $\mathbb{V}$ is isomorphic to one of the following algebras:*

1. *the algebra $\mathbb{L}^n$ (Example 2.2),*
2. *the algebra $\mathbb{S}^n$ of all $n \times n$ real symmetric matrices (Example 2.1),*
3. *the algebra $\mathbb{H}^n$ of all $n \times n$ complex Hermitian matrices with $X \circ Y = \frac{1}{2}(XY + YX)$,*
4. *the algebra $\mathbb{Q}^n$ of all $n \times n$ quaternion Hermitian matrices with $X \circ Y = \frac{1}{2}(XY + YX)$,*
5. *the algebra $\mathbb{O}^3$ of all 3×3 octonion Hermitian matrices with $X \circ Y = \frac{1}{2}(XY + YX)$.*

The first four algebras are special, and the last one is an exceptional Jordan algebra of dimension 27.

Proof. For the first part of the results, see proofs of Corollary IV.1.5 and Theorem V.3.7 in [23]. The first four algebra are special. A proof can be found either in [68, p. 49] or in [89, Section 4]. The last statement comes from the Albert's Exceptional Theorem [1]. □

2.4 Associative Peirce decomposition

In this section, we present some decomposition structures in associative algebras. They extend the following formula (2.5) of a square matrix in the block form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{2.5}$$

where A_{11} is square. Our terminology and notation follow from [68, Part II 8.3]. We apply these structures to deduce formulae of the inverse of an element in terms of the "block" form.

Let c be an idempotent in an associative algebra $\mathbb{A}$. For each $x \in \mathbb{A}$, it has a decomposition

$$x = exe = (c + e - c)x(c + e - c) = cxc + cx(e - c) + (e - c)xc + (e - c)x(e - c).$$

If we set $e_1 = c, e_2 = e - c$, this becomes

$$x = e_1xe_1 + e_1xe_2 + e_2xe_1 + e_2xe_2,$$

and we obtain an *associative Peirce decomposition* of $\mathbb{A}$

$$\mathbb{A} = \mathbb{A}_{11} \oplus \mathbb{A}_{12} \oplus \mathbb{A}_{21} \oplus \mathbb{A}_{22} \tag{2.6}$$

relative to c, where $\mathbb{A}_{ij} = \{e_ixe_j \mid x \in \mathbb{A}\}$. Moreover, these Peirce spaces $\mathbb{A}_{ij}$ satisfy the *associative Peirce multiplication rules*

$$\mathbb{A}_{ij}\mathbb{A}_{kl} \subseteq \delta_{jk}\mathbb{A}_{il}. \tag{2.7}$$

If the associative algebra $\mathbb{A}$ has an involution $*$, and $c^* = c$ is a self-adjoint idempotent, then the associative Peirce spaces satisfy $\mathbb{A}_{ij}^* = \mathbb{A}_{ji}$. The associative Peirce decomposition of its Jordan subalgebras $\mathbb{V}$ of self-adjoint elements is precisely

$$\mathbb{V} = \mathbb{V}(c, 1) \oplus \mathbb{V}(c, \frac{1}{2}) \oplus \mathbb{V}(c, 0)$$

with

$$\mathbb{V}(c,1)=\mathbb{V}_{11},\ \mathbb{V}(c,0)=\mathbb{V}_{22},$$

$$\mathbb{V}(c,\frac{1}{2})=\mathbb{V}_{12}\oplus\mathbb{V}_{12}^{*}=\{x_{12}+x_{12}^{*}\mid x_{12}\in\mathbb{V}_{12}\}.$$

The next result extends the block-matrix inverse formulae.

Proposition 2.11. *Let $\mathbb{A}$ be an associative algebra with the decomposition (2.6), for $x\in\mathbb{A}$, we write $x=x_{11}+x_{12}+x_{21}+x_{22}$ with $x_{ij}\in\mathbb{A}_{ij}, i,j\in\{1,2\}$. If x is invertible and x_{11} is invertible in $\mathbb{A}_{11}$ with inverse x_{11}^{-1}, then*

$$\begin{aligned}
(x^{-1})_{11}&=x_{11}^{-1}+x_{11}^{-1}x_{12}(x_{22}-x_{21}x_{11}^{-1}x_{12})^{-1}x_{21}x_{11}^{-1},\\
(x^{-1})_{12}&=-x_{11}^{-1}x_{12}(x_{22}-x_{21}x_{11}^{-1}x_{12})^{-1},\\
(x^{-1})_{21}&=-(x_{22}-x_{21}x_{11}^{-1}x_{12})^{-1}x_{21}x_{11}^{-1},\\
(x^{-1})_{22}&=(x_{22}-x_{21}x_{11}^{-1}x_{12})^{-1}.
\end{aligned}\tag{2.8}$$

If x is invertible and x_{22} is invertible in $\mathbb{A}_{22}$ with inverse x_{22}^{-1}, then

$$\begin{aligned}
(x^{-1})_{11}&=(x_{11}-x_{12}x_{22}^{-1}x_{21})^{-1},\\
(x^{-1})_{12}&=-(x_{11}-x_{12}x_{22}^{-1}x_{21})^{-1}x_{12}x_{22}^{-1},\\
(x^{-1})_{21}&=-x_{22}^{-1}x_{21}(x_{11}-x_{12}x_{22}^{-1}x_{21})^{-1},\\
(x^{-1})_{22}&=x_{22}^{-1}+x_{22}^{-1}x_{21}(x_{11}-x_{12}x_{22}^{-1}x_{21})^{-1}x_{12}x_{22}^{-1}.
\end{aligned}\tag{2.9}$$

Proof. We now prove the equation (2.8). The second equation is similarly proved. Suppose that $x^{-1}=y=y_{11}+y_{12}+y_{21}+y_{22}$ with $y_{ij}\in\mathbb{A}_{ij}, i,j\in\{1,2\}$, in view of the associative Peirce multiplication rules, it yields

$$\begin{aligned}
e_1+e_2=e=&(x_{11}+x_{12}+x_{21}+x_{22})(y_{11}+y_{12}+y_{21}+y_{22})\\
=&x_{11}y_{11}+x_{11}y_{12}+x_{12}y_{21}+x_{12}y_{22}+x_{21}y_{11}+x_{21}y_{12}+x_{22}y_{21}+x_{22}y_{22},
\end{aligned}$$

whence

$$\begin{aligned}
&x_{11}y_{11}+x_{12}y_{21}=e_1,\\
&x_{11}y_{12}+x_{12}y_{22}=0,
\end{aligned}$$

$$x_{21}y_{11} + x_{22}y_{21} = 0,$$
$$x_{21}y_{12} + x_{22}y_{22} = e_2.$$

If x_{11} is invertible, direct computation yields (2.8). □

Remark 2.2. *Let $\mathbb{V}$ be a Euclidean Jordan algebra derived from the associative algebra $\mathbb{A}$, for any $x \in \mathbb{V}$, we can write $x = x_{11} + x_{12} + x_{12}^* + x_{22}$ as its associative Peirce decomposition relative to c. If x is invertible and x_{11} is invertible in $\mathbb{A}_{11}$ with inverse x_{11}^{-1}, then the Schur complement of x_{11} in x is*

$$x_{22} - x_{12}^* x_{11}^{-1} x_{12}.$$

Likewise, If x is invertible and x_{22} is invertible in $\mathbb{A}_{22}$ with inverse x_{22}^{-1}, then the Schur complement of x_{22} in x is

$$x_{11} - x_{12} x_{22}^{-1} x_{12}^*.$$

Chapter 3

Convex quadratic symmetric cone programs and eigenvalue optimization problems

In this chapter, we will study convex quadratic programming over symmetric cones. The relation between the convex quadratic symmetric cone programming and the maximum eigenvalue problem is explored, that is, the former can be cast as an eigenvalue minimization problem. In addition, this chapter includes some background material on nonsmooth analysis and preliminary results which will be used later.

3.1 Relating convex quadratic symmetric cone programs to eigenvalue optimization problems

In this section, we shall reformulate the dual convex quadratic symmetric cone program (D) as the eigenvalue minimization problem (EigForm) with the help of Assumption **A1** given in the introduction.

We discuss the relationship of the dual convex quadratic symmetric cone program (D) with the eigenvalue minimization problem (EigForm) in the following proposition. The proof of this proposition shares the same argument as that in the proof of Proposition 5.1.1 in [35].

Proposition 3.1. *If Assumption* **A1** *is satisfied, then the dual program (D) is equivalent to the eigenvalue problem (Eigform) for* $\alpha = \max\{0, b^T\bar{y}\}$*. If, in addition, the optimal solution set* $\mathcal{O}_D$ *of the dual problem (D) is nonempty, then the optimal solution set* $\mathcal{O}_\star$ *of the eigenvalue problem (Eigform) is* $\left\{(x^\star, y^\star) + \tau(\mathbf{0}, \bar{y}) \;\middle|\; (x^\star, y^\star) \in \mathcal{O}_D, \tau \in \mathbb{R}\right\}$*. Furthermore, all feasible solutions* x *for the CQSCP (P) satisfy* $tr(x) = \alpha$*.*

Proof. Let $\bar{y}$ satisfy Assumption **A1**, and

$$\mathbf{F} = \{(x, y) + \tau(\mathbf{0}, \bar{y}) \mid (x, y) \in \mathbb{V} \times \mathbb{R}^m, \tau \geq \lambda_1(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x))\}.$$

By virtue of Assumption **A1**, for any $(x, y + \tau\bar{y}) \in \mathbf{F}$,

$$\begin{aligned}
&\mathcal{Q}(x) + \mathcal{A}^*(y + \tau\bar{y}) + \hat{s} \\
&= \tau(\mathcal{A}^*(\bar{y})) + \mathcal{Q}(x) + \mathcal{A}^*(y) + \hat{s} = \tau e - (-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)) \\
&\succeq \lambda_1(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x))e - (-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)) \in \mathbb{K}.
\end{aligned}$$

Hence, any element of $\mathbf{F}$ is feasible for (D).

Now, consider the cases $b^T\bar{y} < 0$ and $b^T\bar{y} \geq 0$.

Case 1: $b^T\bar{y} < 0$.

In this case, $\alpha = 0$. Then $f(x, y) = g(x, y)$ for any $(x, y) \in \mathbb{V} \times \mathbb{R}^m$.

Obviously, each element in $\mathbf{F}$ is also feasible for (Eigform).

For any $(x, y + \tau\bar{y}) \in \mathbf{F}$,

$$g(x, y + \tau\bar{y}) = \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle + b^T(y + \tau\bar{y}) = \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle + b^T y + \tau b^T\bar{y} \to -\infty$$

as $\tau \to +\infty$.

We conclude that (D) is equivalent to (Eigform).

Case 2: $b^T\bar{y} \geq 0$.

In this case, $\alpha = b^T\bar{y}$. We shall show that feasible solutions of (Eigform) are in one to one correspondence to those of (D).

For any fixed $(x, y) \in \mathbb{V}\times\mathbb{R}^m$, we now claim that $f(x, y)$ is constant along directions $\tau(\mathbf{0}, \bar{y})$ for any $\tau \in \mathbb{R}$, i.e.

$$f(x, y + \tau\bar{y}) = f(x, y). \tag{3.1}$$

For any $(u,z) \in \{(x,y) + \tau(\mathbf{0},\bar{y}) \mid \tau \in \mathbb{R}\} \supseteq \mathbf{F}$, by Assumption **A1**, it holds for some $\tau \in \mathbb{R}$ that

$$\begin{aligned} f(u,z) &= f(x, y+\tau\bar{y}) \\ &= \alpha\lambda_1(-\hat{s} - \mathcal{A}^*(y+\tau\bar{y}) - \mathcal{Q}(x)) + b^T(y+\tau\bar{y}) + \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle \\ &= \alpha\lambda_1(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x) - \tau e) + b^T y + \tau b^T\bar{y} + \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle \\ &= \alpha\lambda_1(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)) - \alpha\tau + b^T y + \tau\alpha + \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle = f(x,y). \end{aligned}$$

Setting $\tau_0 = \lambda_1(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x))$, we get

$$(x,y) + \tau_0(\mathbf{0},\bar{y}) \in \mathbf{F} \text{ and } \lambda_1(-\hat{s} - \mathcal{A}^*(y+\tau_0\bar{y}) - \mathcal{Q}(x)) = 0.$$

Thus $(x, y+\tau_0\bar{y})$ is a feasible solution of (D) and

$$g(x, y+\tau_0\bar{y}) = f(x, y+\tau_0\bar{y}). \tag{3.2}$$

Conversely, any feasible solution $(\tilde{x},\tilde{y})$ of (D) is also feasible for (Eigform).

Let $\tilde{\tau} = \lambda_1(-\hat{s} - \mathcal{A}^*(\tilde{y}) - \mathcal{Q}(\tilde{x}))$. Then $\tilde{\tau} \le 0$, and

$$\mathcal{Q}(\tilde{x}) + \mathcal{A}^*(\tilde{y}+\tilde{\tau}\bar{y}) + \hat{s} = \tilde{\tau}e - (-\hat{s} - \mathcal{A}^*(\tilde{y}) - \mathcal{Q}(\tilde{x})) \succeq \mathbf{0}.$$

Thus $(\tilde{x}, \tilde{y}+\tilde{\tau}\bar{y})$ is a feasible solution of (D).

Note that

$$\begin{aligned} g(\tilde{x}, \tilde{y}+\tilde{\tau}\bar{y}) &= \frac{1}{2}\langle \tilde{x}, \mathcal{Q}(\tilde{x})\rangle + b^T(\tilde{y}+\tilde{\tau}\bar{y}) \\ &= \frac{1}{2}\langle \tilde{x}, \mathcal{Q}(\tilde{x})\rangle + b^T\tilde{y} + \alpha\lambda_1(-\hat{s} - \mathcal{A}^*(\tilde{y}) - \mathcal{Q}(\tilde{x})), \end{aligned}$$

namely,

$$g(\tilde{x}, \tilde{y}+\tilde{\tau}\bar{y}) = f(\tilde{x},\tilde{y}). \tag{3.3}$$

Moreover,

$$g(\tilde{x}, \tilde{y}+\tilde{\tau}\bar{y}) \le \frac{1}{2}\langle \tilde{x}, \mathcal{Q}(\tilde{x})\rangle + b^T\tilde{y} = g(\tilde{x},\tilde{y}). \tag{3.4}$$

Recall that

$$\mathcal{O}_D = arg\min\left\{g(x,y) \;\middle|\; \mathcal{Q}(x) + \mathcal{A}^*(y) + \hat{s} \in \mathbb{K}\right\},$$
$$\mathcal{O}_\star = arg\min\left\{f(x,y) \;\middle|\; (x,y) \in \mathbb{V} \times \mathbb{R}^m\right\}.$$

In the subsequent analysis, we shall show that

$$\mathcal{O}_\star = \left\{(x^\star, y^\star) + \tau(\mathbf{0}, \bar{y}) \;\middle|\; (x^\star, y^\star) \in \mathcal{O}_D, \tau \in \mathbb{R}\right\}.$$

In fact, suppose that $(x^\star, y^\star) \in \mathcal{O}_\star$, then $(x^\star, y^\star + \tau^\star \bar{y}) \in \mathcal{O}_\star$ by (3.1), where

$$\tau^\star = \lambda_1(-\hat{s} - \mathcal{A}^*(y^\star) - \mathcal{Q}(x^\star)).$$

We then claim that $(x^\star, y^\star + \tau^\star \bar{y}) \in \mathcal{O}_D$. Otherwise, there exists some $(\hat{x}, \hat{y})$ that is feasible for (D), but $g(\hat{x}, \hat{y}) < g(x^\star, y^\star + \tau^\star \bar{y})$. Let $\hat{\tau} = \lambda_1(-\hat{s} - \mathcal{A}^*(\hat{y}) - \mathcal{Q}(\hat{x}))$. We have

$$f(\hat{x}, \hat{y}) \overset{(3.3)}{=} g(\hat{x}, \hat{y} + \hat{\tau}\bar{y}) \overset{(3.4)}{\leq} g(\hat{x}, \hat{y}) < g(x^\star, y^\star + \tau^\star \bar{y}) \overset{(3.2)}{=} f(x^\star, y^\star + \tau^\star \bar{y}) \overset{(3.1)}{=} f(x^\star, y^\star),$$

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

Similarly, we can show that if $(x^\star, y^\star) \in \mathcal{O}_D$, then $(x^\star, y^\star) \in \mathcal{O}_\star$. Together with (3.1), it then follows $(x^\star, y^\star) + \tau(\mathbf{0}, \bar{y}) \in \mathcal{O}_\star$ for any $\tau \in \mathbb{R}$.

Therefore, for any $(x^\star, y^\star) \in \mathcal{O}_D$,

$$g(x^\star, y^\star) \geq g(x^\star, y^\star + \tau^\star \bar{y}) = f(x^\star, y^\star),$$

whence $g(x^\star, y^\star) = f(x^\star, y^\star)$.

Consequently, we prove that (D) is equivalent to (Eigform).

If x is a feasible solution of (P), then $\mathcal{A}(x) = b$. Thus,

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

Hence $tr(x) = b^T \bar{y} = \alpha$. □

Remark 3.1. *Assumption* **A1** *has been used in [35, Chapter 5]. In this thesis, we are* *just concerned with convex quadratic symmetric cone programs whose primal feasible set is bounded. By [35, Corollary 5.1.3], such programs satisfy Assumption* **A1**. *In fact, almost all real-life problems have bounded feasible regions. For instance, semidefinite programs arising from semidefinite relaxations of quadratic* $0-1$ *programming, graph coloring and graph partitioning, see [37, Chapter 11] and references therein, satisfy this property. In addition, we shall see that the nearest correlation matrix problem also satisfies Assumption* **A1**, *see Example 3.1.*

Example 3.1. *Now we consider the nearest correlation matrix problem. The problem (P) resulting from Problem (1.1) has* $\mathcal{Q} = \mathcal{L}^2$, $\hat{s} = -\mathcal{L}^2(B)$, $b = \mathbf{1} \in \mathbb{R}^m$ *and* $\mathcal{A}(X) = (\langle A_1, X\rangle, \cdots, \langle A_n, X\rangle)^T$, *where* $A_i = \mathbf{1}_i \mathbf{1}_i^T$, $i = 1, \cdots, n$. *Thus* $\mathcal{A}^*(y) = Diag(y)$ *for any* $y \in \mathbb{R}^n$. *Moreover Problem (1.1) satisfies Assumption* **A1**. *Indeed,* I *is strictly feasible for Problem (1.1) and* $\mathcal{A}^*(\mathbf{1}) = I$. *By Proposition 3.1, the equivalent form of (Eigform) resulting from Problem (1.1) is*

$$\min_{X,y} \quad n\lambda_1(\mathcal{L}^2(B) - Diag(y) - \mathcal{L}^2(X)) + \mathbf{1}^T y + \frac{1}{2}\langle X, \mathcal{L}^2(X)\rangle. \tag{3.5}$$

This example shows that Assumption **A1** *is satisfied in real applications.*

By virtue of Proposition 3.1, we see that, in the case that $\alpha = 0$, (Eigform) is trivial: for any $y \in \mathbb{R}^m$, $(\mathbf{0}, y)$ is optimal if $b = \mathbf{0}$, otherwise it is unbounded. Henceforth, we suppose that $\alpha > 0$.

To solve the eigenvalue minimization problem (Eigform), the main difficulty arises from the nondifferentiability of $f(x, y)$. In order to overcome this difficulty, one can resort to nonsmooth optimization techniques. Since a minimum point $(x^\star, y^\star)$ is characterized by $(\mathbf{0}, \mathbf{0}) \in \partial f(x^\star, y^\star)$, we are interested in the problem of computing the subdifferential, which will follow in the next section.

3.2 Subdifferential and approximate subdifferential of the largest eigenvalue function

Recall that the maximum eigenvalue function $\lambda_1 : \mathbb{V} \to \mathbb{R}$ is convex and not differentiable in general, see, e.g., [80]. Hence, minimizing the maximum eigenvalue is a nonsmooth optimization problem. This section includes some background material on nonsmooth analysis and preliminary results on the subgradient (see, e.g., [85, Section 12]) and ϵ-subgradient (see, e.g., [42, Chapter XI]) which will be used later.

The following result gives the calculus rules on the subdifferential and approximate subdifferential.

Theorem 3.1. *For any* $(x, y) \in \mathbb{V} \times \mathbb{R}^m$ *and* $\epsilon > 0$, *there holds*

$$\partial f(x, y) = \left\{ (\mathcal{Q}(x - w), b - \mathcal{A}(w)) \in \mathbb{V} \times \mathbb{R}^m \;\middle|\; w \in \alpha \partial \lambda_1(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)) \right\}, \tag{3.6}$$

$$\partial_\epsilon f(x, y) \supseteq \left\{ (\mathcal{Q}(x - w), b - \mathcal{A}(w)) \in \mathbb{V} \times \mathbb{R}^m \;\middle|\; w \in \alpha \partial_\epsilon \lambda_1(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)) \right\}. \tag{3.7}$$

Proof. We set $f_1(x, y) := \alpha \lambda_1(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x))$ and $f_2(x, y) := b^T y + \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle$, then $f(x, y) = f_1(x, y) + f_2(x, y)$. Clearly, $\nabla f_2(x, y) = (\mathcal{Q}(x), b)$. Applying a classical chain rule [42, XI. Theorem 3.2.1], we obtain

$$\partial_\epsilon f_1(x, y) = \left\{ -(\mathcal{Q}(w), \mathcal{A}(w)) \;\middle|\; w \in \alpha \partial_\epsilon \lambda_1(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)) \right\} \text{ for } \epsilon \geq 0,$$

where we use the convention $\partial_0 f = \partial f$. Therefore, (3.6) and (3.7) follow from [42, XI. Corollary 3.1.2] and [42, XI. Theorem 3.1.1], respectively. □

From Theorem 3.1, we see that characterizations of the maximum eigenvalue of $v \in \mathbb{V}$ are useful. We are now in a position to present some ingredients for the nonsmooth analysis of the largest eigenvalue.

We set

$$\mathcal{W} = \{w \in \mathbb{K} \mid tr(w) = 1\}.$$

The next result is a generalization to Euclidean Jordan algebras of a well-known variational description of the largest eigenvalue of $v \in \mathbb{V}$.

Theorem 3.2. *Let* $\mathbb{V}$ *be a Euclidean Jordan algebra with rank* $r \geq 2$. *For any* $v \in \mathbb{V}$, *we obtain*

$$\lambda_1(v) = \max_{w \in \mathcal{W}} \langle v, w\rangle. \tag{3.8}$$

Proof. See Lemma 20 of [6]. □

The following proposition characterizes the subdifferential of $\lambda_1(\cdot)$ at $v \in \mathbb{V}$.

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

Proof. It is straightforward by Proposition 33 in [6], Proposition 2.9 and (iv) of Proposition 2.7. □

In many practical problems, exact eigenvalues and the corresponding idempotents are unavailable, as usually the exact values are irrational; then they are usually computed by means of some iterative algorithm. Therefore it is more reasonable to consider the approximate subdifferential of the maximum eigenvalue function of $v \in \mathbb{V}$, which is the ϵ-subdifferential and visualized by the following proposition.

Proposition 3.3. *For any $\epsilon \geq 0$, it holds*

$$\partial_\epsilon \lambda_1(v) = \left\{u \in \mathcal{W} \middle| \langle u, v\rangle \geq \;\lambda_1(v) - \epsilon\right\}. \tag{3.9}$$

Proof. In view of Proposition 1.1 in [41], we have

$$u \in \partial_\epsilon \lambda_1(v) \text{ if and only if } \;\lambda_1(v) + \lambda_1^\sharp(u) - \langle v, u\rangle \leq \epsilon,$$

where $\lambda_1^\sharp(\cdot)$ is the conjugate of $\lambda_1(\cdot)$. Noting that (3.8) implies that $\lambda_1(\cdot)$ is the support function of the closed covex set $\mathcal{W}$, by Theorem 13.2 in [85], we obtain that $\lambda_1^\sharp(\cdot)$ is the indicator function of $\mathcal{W}$. The proof is complete. □

For later use, we introduce the following concept. For an element $x \in \mathbb{V}$, a triplet $(\lambda_{ap}, \tilde{c}, \epsilon)$ is said to be a *Ritz-type triplet* for $\lambda_1(x)$ if

$$\tilde{c} \text{ is a primitive idempotent,} \tag{3.10}$$

$$\lambda_{ap} = tr(x \circ \tilde{c}), \epsilon = \|x \circ \tilde{c} - \lambda_{ap}\tilde{c}\|_F, \tag{3.11}$$

$$\begin{aligned} &|\lambda_i(x) - \lambda_{ap}| \geq |\lambda_1(x) - \lambda_{ap}| \quad \text{and} \\ &\left|\frac{\lambda_i(x) + \lambda_j(x)}{2} - \lambda_{ap}\right| \geq |\lambda_1(x) - \lambda_{ap}|,\ 1 \leq i < j \leq r. \end{aligned} \tag{3.12}$$

For any $x \in \mathbb{V}$, if there exist a scalar λ and an idempotent c such that

$$x \circ c = \lambda c,$$

we say that c is an idempotent corresponding to the eigenvalue λ of x. We will see that a *Ritz-type triplet* depicts a pair of approximate eigenvalues and the corresponding idempotents for an element of $\mathbb{V}$.

Proposition 3.4. *Let $v \in \mathbb{V}$ and $\epsilon \geq 0$. If $(\lambda_{ap}, \tilde{c}, \epsilon)$ is a* Ritz-type triplet *for $\lambda_1(v)$, then*

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

Proof. Let $v = \sum_{i=1}^r \lambda_i(v) c_i$ be a spectral decomposition of v, and $\tilde{c} = \sum_{i=1}^r x_i c_i + \sum_{i<j} x_{ij}$ be the Pierce decomposition of $\tilde{c}$ with respect to the Jordan frame $\{c_1, \cdots, c_r\}$. From Proposition 2.6, it holds $\lambda_{ap} = tr(v \circ \tilde{c}) \leq \lambda_1(v)$. Since $\tilde{c}$ is a primitive idempotent, $\|\tilde{c}\|_F^2 = \sum_{i=1}^r x_i^2 + \sum_{i<j} \|x_{ij}\|_F^2 = 1$.

On the other hand, we have

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

Therefore, we obtain

$$\begin{aligned}
\epsilon = \|v \circ \tilde{c} - \lambda_{ap}\tilde{c}\|_F &= \sqrt{\sum_{i=1}^r \big[\lambda_i(v) - \lambda_{ap}\big]^2 x_i^2 + \sum_{i<j} \left[\frac{\lambda_i(v) + \lambda_j(v)}{2} - \lambda_{ap}\right]^2 \|x_{ij}\|_F^2} \\
&\geq \sqrt{\sum_{i=1}^r \big[\lambda_1(v) - \lambda_{ap}\big]^2 x_i^2 + \sum_{i<j} \big[\lambda_1(v) - \lambda_{ap}\big]^2 \|x_{ij}\|_F^2} \\
&= |\lambda_1 - \lambda_{ap}| \sqrt{\sum_{i=1}^r x_i^2 + \sum_{i<j} \|x_{ij}\|_F^2} = \lambda_1 - \lambda_{ap},
\end{aligned}$$

where the inequality follows from (3.12).

Therefore $\lambda_{ap} \leq \lambda_1(v) \leq \lambda_{ap} + \epsilon$.

By Proposition 3.3, it is easy to see that $\tilde{c} \in \partial_\epsilon \lambda_1(v)$. □

Throughout this thesis, corresponding to a *Ritz-type triplet* $(\lambda_{ap}, \tilde{c}, \epsilon)$ for $\lambda_1(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x))$, we define an inexact computation of $f(x, y)$ by

$$f_l(x, y) = \alpha\lambda_{ap} + b^T y + \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle \tag{3.13}$$

and

$$f_u(x, y) = f_l(x, y) + \alpha\epsilon. \tag{3.14}$$

Obviously, $f_l(x, y) \le f(x, y) \le f_u(x, y)$ for any $(x, y) \in \mathbb{V} \times \mathbb{R}^m$.

Chapter 4

An inexact spectral bundle algorithm and its convergence analysis

The aim of this chapter is to propose an inexact spectral bundle algorithm for solving the largest eigenvalue optimization problem (EigForm), which is derived from the spectral bundle method [36, 38]. It is "inexact" in the sense that it does not require to compute the exact eigenvalues. Instead, the approximated eigenvalues satisfy a sequence of prescribed tolerances δ^k, which is a decreasing sequence that tends to zero. Such management is to ensure convergence. The proposed Algorithm 4.4 finds two sequences $\{(x^k, y^k)\}$ and $\{w^k\}$ simultaneously, whose limit points are optimal solutions for the eigenvalue minimization problem (EigForm) and the CQSCP (P), respectively.

In this chapter, we first review some general theoretical issues related to bundle methods in Section 4.1. After some preparations in Sections 4.2.1–4.2.4, the algorithm is described in Section 4.2.5. In particular, we design an orthonormalization process to construct an approximate model of the convex set $\alpha\mathcal{W}$ in Section 4.2.3. We provide convergence analysis in Section 4.3.

4.1 Bundle methods

In this section, we briefly recall the bundle method to solve the problem of minimizing a nondifferentiable and convex function $f(x)$ on $\mathbb{R}^n$, i.e.

$$\min_{x\in\mathbb{R}^n} f(x).$$

The bundle method is one of the most widely studied methods for solving nonsmooth optimization problems. It is a first-order method. It combines both descent and stability properties. The bundle method can be interpreted as a cutting plane algorithm stabilized by a quadratic penalty. This introduction to bundle methods is inspired by the book [11].

4.1.1 Cutting-plane method

Essentially, the cutting-plane method (see, e.g., [11, section 9.3.2]) generates a sequence of iterates, starting from $x^1 \in S$ (S is a given compact convex set containing a minimum point of f) and defined by

$$x^{k+1} = arg \min_{y\in S} \check{f}_k(y),$$

where

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

is a piecewise linear approximation to f, and $g(x)$ is a subgradient of f at the point x.

For general convex functions, the cutting-plane algorithm may encounter several serious problems as follows.

1) The sequence does not necessarily have decreasing objective value, see paragraph immediately after [11, Remark 9.5].

2) It may present instabilities and bad numerical behavior. For example, when “horizontal” affine functions are added to the model, iterates would move far away from the set of minima and increase the current function value [11, Example 9.7].

3) As k approaches infinity, there would be infinite many affine functions being

added to the model, and results in that $\check{f}_k(y)$ becomes extremely difficult to compute [11, Remark 9.8].

We will see that bundle methods can overcome these drawbacks.

4.1.2 General bundle method

The bundle method (see, e.g., [11, Section 10]) generates a sequence of *trial points* $\{y^k\}_{k=1}^{\infty}$, which contains a subsequence of *stability centers* $\{x^k\}_{k=1}^{\infty}$, starting from y^0 and defined by

$$y^{k+1} = arg \min_{y \in \mathbb{R}^n} \{\varphi_k(y) + ||y - x^k||_k^2\}, \tag{4.1}$$

where φ_k is an underestimate of f, and the norm $\|\cdot\|_k$ is to measure distance to x^k and is chosen to prevent big oscillations at each iteration.

Remark 4.1. *1.Trial points $\{x^i\}_{i=1}^k$are used to define a model φ_k approximating the objective function f. For instance, for all $x^i, i = 1, \dots, k$, we find values $f(x^i)$ and subgradients $g(x^i)$ of f that product a piecewise linear approximation $\check{f}_k$ to f, which can be a choice of φ_k.*

2. Stability centers mark sufficient decrease in the value of the objective function f.

The choice of a stability center x^k should ensure that $f(x^k)$ is the "best" value obtained so far and the stability center x^k decreases the value of f. The quality of trial points is measured using the nominal decrease ε^k. Only "good" trial points, i.e. those satisfying the relation

$$f(y^{k+1}) \leq f(x^k) - m\varepsilon^k \tag{4.2}$$

for $m \in (0, 1)$, will become stability centers.

Suppose that the following oracle is available.

Oracle 4.1. *For any given point x, it returns the value of the function $f(x)$ and only one arbitrary subgradient $g(x) \in \partial f(x)$.*

We are now in a position to go on at length about the bundle method.

Algorithm 4.1. *[11, Algorithm 10.1] Let tol* ≥ 0 *and* $m \in (0,1)$ *be given parameters. Choose* x^0*, call* **Oracle 1** *with* $x = x^0$*, construct the model* φ_0*, and set* $||\cdot||_0$*. Set* $k = 0$*,* $\varepsilon^0 = \infty$ *and* $y^0 = x^0$*.*

1. ***(Implementable stopping test)*** *If* $\varepsilon^k \leq tol$*, stop.*

2. ***(Trial point)*** *Compute* y^{k+1} *from (4.1). Define* $\varepsilon^{k+1} \geq 0$*.*

3. ***(Assessing the trial point)*** *Call* **Oracle 1** *with* $x = y^{k+1}$*.*

   Descent test:

$$f(x^k) - f(y^{k+1}) \geq m\varepsilon^k ? \begin{cases} Yes: & x^{k+1} := y^{k+1} (\textit{descent-step}) \\ No: & x^{k+1} := x^k (\textit{null-step}) \end{cases}$$

4. ***(Improving the model)*** *Append* y^{k+1} *to the model, i.e. construct* φ_{k+1}*. Define* $||\cdot||_{k+1}$ *for the next iteration. Set* $k := k+1$*, go to step 1.*

4.2 An inexact spectral bundle algorithm

The aim of this section is to introduce the main ingredients of our proposed method. With the help of these ingredients, we present an inexact spectral bundle algorithm, which is a realization of the general scheme for bundle methods.

4.2.1 The computation of trial points

In this subsection, we first define an approximate of the cost function f, which is different from the cutting plane model of f. We will then use such approximate model stabilized by a quadratic penalty to specify a sequence of minimization subproblems for determining trial points.

For present purposes we restrict ourselves to the current iterate k. Suppose that $c_{1k}, \cdots, c_{r_k k}$ are mutually orthogonal primitive idempotents. We denote

$$e^k_{r_k} = c_{1k} + \cdots + c_{r_k k},$$
$$\mathbb{V}^{(r_k)}_k = \{x \in \mathbb{V} \mid x \circ e^k_{r_k} = x\},$$

$$\mathbb{K}^{(r_k)} = \mathbb{V}_k^{(r_k)} \cap \mathbb{K}.$$

Inspired by the characterization of subdifferential given by Proposition 3.2, we use

$$\widehat{\mathcal{W}}^k = \left\{ v + \xi \bar{w}_k \;\middle|\; tr(v) + \xi = \alpha, v \in \mathbb{K}^{(r_k)}, \xi \geq 0 \right\} \tag{4.3}$$

as an approximation of $\alpha\mathcal{W}$ with $\widehat{\mathcal{W}}^k \subseteq \alpha\mathcal{W}$, where $\bar{w}_k \in \mathcal{W}$ is called the *aggregate subgradient*. A special advantage of this approximation is that it can keep the rank r_k of $\mathbb{V}_k^{(r_k)}$ small.

Note that $\frac{1}{2}\langle x, \mathcal{Q}(x)\rangle = \frac{1}{2}\langle \hat{x}^k, \mathcal{Q}(\hat{x}^k)\rangle + \langle x - \hat{x}^k, \mathcal{Q}(\hat{x}^k)\rangle + \frac{1}{2}\langle x - \hat{x}^k, \mathcal{Q}(x - \hat{x}^k)\rangle$. For any given $w \in \alpha\mathcal{W}$, let an *approximate partial linearization* of f be

$$\begin{aligned} f_w(x,y) &= \langle -\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x), w\rangle + b^T y + \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle \\ &= \langle -\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x), w\rangle + b^T y + \frac{1}{2}\langle \hat{x}^k, \mathcal{Q}(\hat{x}^k)\rangle + \langle x - \hat{x}^k, \mathcal{Q}(\hat{x}^k)\rangle \\ &+ \frac{1}{2}\langle x - \hat{x}^k, \mathcal{Q}(x - \hat{x}^k)\rangle \end{aligned} \tag{4.4}$$

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

$$\max_{i=1,\ldots,k} \left\{ \bar{f}_i(x,y) := f(x^i, y^i) + \langle \mathcal{Q}(x^i - c_i), x - x^i\rangle + \langle b - \mathcal{A}(c_i), y - y^i\rangle \right\},$$

where c_i is a primitive idempotent corresponding to $\lambda_1(-\hat{s} - \mathcal{A}^*(y^i) - \mathcal{Q}(x^i))$ for $i = 1, \cdots, k$. In effect, by setting $\widehat{\mathcal{W}}^k = conv\{c_i \mid i = 1, \cdots, k\}$, we have

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

However, we see that the model $f_{\widehat{\mathcal{W}}^k}(x,y)$ coincides with the cutting plane model for (linear) semidefinite programming, see [37, p. 314].

Remark 4.2. *The function $f_{\widehat{\mathcal{W}}^k}(x,y)$ is a minorant of $f(x,y)$ for any $(x,y) \in \mathbb{V} \times \mathbb{R}^m$.*

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

$$L^k(x, y, w) = f_w(x, y) + \frac{\nu_k}{2}(\|x - \hat{x}^k\|^2_{\mathcal{M}_{x,k}} + \|y - \hat{y}^k\|^2_{\mathbf{M}_{y,k}}). \tag{4.5}$$

Recall that $\nu_k,\ \mathcal{M}_{x,k}$ and $\mathbf{M}_{y,k}$ have been specified in (1.2).

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

$$\phi^k(x, y) = \max_{w \in \widehat{\mathcal{W}}^k} L^k(x, y, w) = f_{\widehat{\mathcal{W}}^k}(x, y) + \frac{\nu_k}{2}(\|x - \hat{x}^k\|^2_{\mathcal{M}_{x,k}} + \|y - \hat{y}^k\|^2_{\mathbf{M}_{y,k}}). \tag{4.6}$$

In order to determine properties of $\phi^k(x, y)$, we need the following two propositions.

Proposition 4.1. *We can rewrite* $f_{\widehat{\mathcal{W}}^k}(x, y)$ *as*

$$\alpha \max \left\{ \lambda_1(P_{e^k_{r_k}}(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x))), \langle -\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x), \bar{w}_k \rangle \right\} + b^T y + \frac{1}{2}\langle x, \mathcal{Q}(x) \rangle.$$

Proof. For any fixed (x, y), $f_{\widehat{\mathcal{W}}^k}(x, y)$ is a convex program which maximizes a linear function over a compact convex set, thus the maximum is equal to the maximum over the extreme points of the set $\widehat{\mathcal{W}}^k$.

Since $\alpha \bar{w}_k \in \widehat{\mathcal{W}}^k$ and $\left\{ u \in \mathbb{V} \;\middle|\; tr(u) = \alpha, u \in \mathbb{K}^{(r_k)} \right\} =: \mathcal{W}_0 \subseteq \widehat{\mathcal{W}}^k$, it holds

$$conv(\{\alpha \bar{w}_k\} \cup \mathcal{W}_0) \subseteq \widehat{\mathcal{W}}^k.$$

Conversely, for any $w = v + \xi \bar{w}_k \in \widehat{\mathcal{W}}^k$ with $0 < \xi < \alpha$, and $u = \frac{\alpha}{\alpha - \xi} v$, it holds $tr(u) = \frac{\alpha}{\alpha - \xi} tr(v) = \alpha$. Thus $conv(\{\alpha \bar{w}_k\} \cup \mathcal{W}_0) = \widehat{\mathcal{W}}^k$. Therefore,

$$f_{\widehat{\mathcal{W}}^k}(x, y) = \max \{ f_{\alpha \bar{w}_k}(x, y), f_{\mathcal{W}_0}(x, y) \}.$$

In light of (iii) and (iv) in Proposition 2.9, we get

$$\begin{aligned} & f_{\mathcal{W}_0}(x, y) \\ = & \max_{w \in \mathcal{W}_0} \left\{ \langle -\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x), P_{e^k_{r_k}}(w) \rangle + b^T y + \frac{1}{2}\langle x, \mathcal{Q}(x) \rangle \right\} \end{aligned}$$

$$= \max_{w \in \mathcal{W}_0} \left\langle P_{e^k_{r_k}}(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)), w \right\rangle + b^T y + \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle,$$

where the second equality comes from the self-adjointness of $P_{e^k_{r_k}}$.

On the other hand, we have

$$\begin{aligned}
&\lambda_1(P_{e^k_{r_k}}(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x))) \\
=&\max_{w \in \mathcal{W}} \left\langle w, P_{(e^k_{r_k})^2}(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)) \right\rangle && \text{(Theorem 3.2)} \\
=&\max_{w \in \mathcal{W}} \left\langle P_{e^k_{r_k}}(w), P_{e^k_{r_k}}(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)) \right\rangle && \text{((iii) of Proposition 2.2)} \\
=&\frac{1}{\alpha} \max_{w \in \mathcal{W}_0} \left\langle w, P_{e^k_{r_k}}(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)) \right\rangle && \text{((iii) and (iv) of Proposition 2.9).}
\end{aligned}$$

Now we can conclude that

$$f_{\mathcal{W}_0}(x, y) = \alpha\lambda_1(P_{e^k_{r_k}}(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x))) + b^T y + \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle.$$

The proof is complete. □

For later use, for any k, we let

$$\mathcal{T}_k = \mathcal{Q} + \nu_k \mathcal{M}_{x,k},$$

which defines a self-adjoint positive definite linear operator from $\mathbb{V}$ to itself.

The following proposition deals with the question of computing a trial point at each iteration of the algorithm. In essence, we obtain new trial point via solving the minimization subproblem $\min_{x,y} \phi^k(x, y)$.

Proposition 4.2. *The minimizer of $\phi^k(x, y)$ over $\mathbb{V} \times \mathbb{R}^m$ possesses the following 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

$$x^{k+1} = \hat{x}^k + \mathcal{T}_k^{-1}(\mathcal{Q}(w^{k+1} - \hat{x}^k)) \quad \text{and} \quad y^{k+1} = \hat{y}^k + \frac{1}{\nu_k}\mathbf{M}_{y,k}^{-1}(\mathcal{A}(w^{k+1}) - b) \tag{4.7}$$

are unique, and w^{k+1} *is an optimal solution of*

$$\begin{aligned}
\min \quad & \tfrac{1}{2}\big\|\mathcal{Q}(w-\hat{x}^k)\big\|^2_{\mathcal{T}_k^{-1}} + \tfrac{1}{2\nu_k}\big\|b-\mathcal{A}(w)\big\|^2_{\mathbf{M}_{y,k}^{-1}} - b^T\hat{y}^k \\
& +\langle \hat{s} + \mathcal{A}^*(\hat{y}^k) + \mathcal{Q}(\hat{x}^k), w\rangle - \tfrac{1}{2}\langle \hat{x}^k, \mathcal{Q}(\hat{x}^k)\rangle \\
\text{s.t.} \quad & w = v + \xi \bar{w}_k \\
& tr(v) + \xi = \alpha \\
& v \in \mathbb{K}^{(r_k)}, \xi \geq 0.
\end{aligned} \tag{SQSCP}$$

Consequently, we have

$$w^{k+1} \in arg \max_{w\in\widehat{\mathcal{W}}^k} L^k(x^{k+1}, y^{k+1}, w), \tag{4.8}$$

and

$$w^{k+1} \in arg \max_{w\in\widehat{\mathcal{W}}^k} f_w(x^{k+1}, y^{k+1}). \tag{4.9}$$

Proof. By virtue of Proposition 4.1, $f_{\widehat{\mathcal{W}}^k}(x,y)$ is equivalent to

$$\begin{aligned}
\min \quad & \alpha\tau + b^T y + \tfrac{1}{2}\langle x, \mathcal{Q}(x)\rangle \\
\text{s.t.} \quad & P_{e^k_{r_k}}(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)) \preceq \tau e^k_{r_k} \\
& \langle -\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x), \bar{w}_k\rangle \leq \tau.
\end{aligned}$$

Thus, from (4.6), $\min\limits_{x,y} \phi^k(x,y)$ can be reformulated as

$$\begin{aligned}
\min \quad & \alpha\tau + b^T y + \tfrac{1}{2}\langle x, \mathcal{Q}(x)\rangle + \tfrac{\nu_k}{2}(\|x-\hat{x}^k\|^2_{\mathcal{M}_{x,k}} + \|y-\hat{y}^k\|^2_{\mathbf{M}_{y,k}}) \\
\text{s.t.} \quad & \tau e^k_{r_k} + P_{e^k_{r_k}}(\mathcal{A}^*(y) + \mathcal{Q}(x) + \hat{s}) \succeq \mathbf{0} \\
& \tau + \langle \mathcal{A}^*(y) + \mathcal{Q}(x) + \hat{s}, \bar{w}_k\rangle \geq 0.
\end{aligned}$$

We define the Lagrangian of the above problem by

$$\begin{aligned}
& L(x,y,\tau,v,\xi) \\
=& \alpha\tau + b^T y + \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle + \frac{\nu_k}{2}(\|x-\hat{x}^k\|^2_{\mathcal{M}_{x,k}} + \|y-\hat{y}^k\|^2_{\mathbf{M}_{y,k}}) \\
& - \langle v, \tau e^k_{r_k} + P_{e^k_{r_k}}(\mathcal{A}^*(y) + \mathcal{Q}(x) + \hat{s})\rangle - \xi(\tau + \langle \mathcal{A}^*(y) + \mathcal{Q}(x) + \hat{s}, \bar{w}_k\rangle),
\end{aligned}$$

where $v \in \mathbb{K}^{(r_k)}$ and $\xi \geq 0$.

By virtue of Assumption **A1**, the above minimization problem is strictly feasible. Moreover, the *Moreau-Yosida regularization* minimization subproblem $\min_{x,y} \phi^k(x,y)$ is finite. Therefore, by the Lagrange Duality theorem, a direct computation implies that there exist $v \succeq \mathbf{0}$ and $\xi \geq 0$ such that

$$
\begin{aligned}
\frac{\partial L}{\partial x} &= \mathcal{Q}(x) - \mathcal{Q}(v + \xi \bar{w}_k) + \nu_k \mathcal{M}_{x,k}(x - \hat{x}^k) = \mathbf{0}, \\
\frac{\partial L}{\partial y} &= b - \mathcal{A}(v + \xi \bar{w}_k) + \nu_k \mathbf{M}_{y,k}(y - \hat{y}^k) = \mathbf{0}, \\
\frac{\partial L}{\partial \tau} &= \alpha - \langle v, e^k_{r_k} \rangle - \xi = 0. \\
& v \in \mathbb{K}^{(r_k)},\ \xi \geq 0
\end{aligned}
$$

Setting $w = v + \xi \bar{w}_k$, it follows

$$
x = \hat{x}^k + \mathcal{T}_k^{-1}(\mathcal{Q}(w - \hat{x}^k)) =: x^k_{min}(w), \tag{4.10}
$$

$$
y = \hat{y}^k + \frac{1}{\nu_k}\mathbf{M}_{y,k}^{-1}(\mathcal{A}(w) - b) =: y^k_{min}(w), \tag{4.11}
$$

Moreover, such (x, y) is unique since $\phi^k(x, y)$ is strictly convex. As for the solution $(v^\star,\ \xi^\star)$ of the above linear systems, let $w^{k+1} = v^\star + \xi^\star \bar{w}_k$, we get the corresponding (x^{k+1}, y^{k+1}) with the desired expressions.

On the other hand, together with (4.5), we obtain, for any $w \in \widehat{\mathcal{W}}^k$,

$$
\begin{aligned}
&\min_{x,y} L^k(x, y, w) = L^k(x^k_{min}(w), y^k_{min}(w), w) \\
=& -\langle \hat{s}, w \rangle + \langle b - \mathcal{A}(w), \hat{y}^k + \frac{1}{\nu_k}\mathbf{M}_{y,k}^{-1}(\mathcal{A}(w) - b) \rangle + \frac{1}{2\nu_k}\big\| b - \mathcal{A}(w) \big\|^2_{\mathbf{M}_{y,k}^{-1}} \\
& - \langle \mathcal{T}_k^{-1}(\mathcal{Q}(w - \hat{x}^k)), \mathcal{Q}(w - \hat{x}^k) \rangle + \frac{1}{2}\langle \mathcal{T}_k^{-1}(\mathcal{Q}(w - \hat{x}^k)), \mathcal{Q}(\mathcal{T}_k^{-1}(\mathcal{Q}(w - \hat{x}^k))) \rangle \\
& + \frac{\nu_k}{2}\big\| \mathcal{T}_k^{-1}(\mathcal{Q}(w - \hat{x}^k)) \big\|^2_{\mathcal{M}_{x,k}} + \frac{1}{2}\langle \mathcal{Q}(\hat{x}^k), \hat{x}^k \rangle \\
=& -\frac{1}{2}\big\| \mathcal{Q}(w - \hat{x}^k) \big\|^2_{\mathcal{T}_k^{-1}} - \frac{1}{2\nu_k}\big\| b - \mathcal{A}(w) \big\|^2_{\mathbf{M}_{y,k}^{-1}} + b^T \hat{y}^k \\
& - \langle \hat{s} + \mathcal{A}^*(\hat{y}^k) + \mathcal{Q}(\hat{x}^k), w \rangle + \frac{1}{2}\langle \hat{x}^k, \mathcal{Q}(\hat{x}^k) \rangle,
\end{aligned}
$$

where the first equality comes from the first order optimal necessary and sufficient condition in unconstrained minimization.

Since $\widehat{\mathcal{W}}^k$ is bounded, by convex minimax duality,

$$
\begin{aligned}
\min_{x,y} \phi^k(x,y) &= \min_{x,y} \max_{w\in\widehat{\mathcal{W}}^k} L^k(x,y,w) \\
&= \max_{w\in\widehat{\mathcal{W}}^k} \min_{x,y} L^k(x,y,w) \\
&= \max_{w\in\widehat{\mathcal{W}}^k} L^k(x^k_{min}(w), y^k_{min}(w), w).
\end{aligned}
$$

Switching the optimization direction in $\max_{w\in\widehat{\mathcal{W}}^k}$ and the sign in $L^k(x^k_{min}(w), y^k_{min}(w), w)$, we get (SQSCP). □

Remark 4.3. *It is easy to see that the subproblem (SQSCP) is an instance of convex quadratic symmetric cone programming. This problem can be solved efficiently by interior point methods (see, for instance, [60]), provided the rank r_k is small, say, around 8.*

4.2.2 The optimality estimates

In this section, we explore the approximate subgradients of the cost function f at a new trial point (x^{k+1}, y^{k+1}) and the current $(\hat{x}^k, \hat{y}^k)$, respectively, which results in the optimality estimates. We obtain a sufficient condition for the point (x^{k+1}, y^{k+1}) or $(\hat{x}^k, \hat{y}^k)$ being optimal. Consequently, it will handle the case of a finite number of iterations.

In light of Proposition 4.2,

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

the necessary and sufficient optimality conditions for the strongly convex problem imply that

$$
\mathbf{0} \in \partial f_{\widehat{\mathcal{W}}^k}(x^{k+1}, y^{k+1}) + \nu_k(\mathcal{M}_{x,k}(x^{k+1}-\hat{x}^k), \mathbf{M}_{y,k}(y^{k+1}-\hat{y}^k))
$$

if and only if there exists some

$$(g_x^k, g_y^k) \in \partial f_{\widehat{\mathcal{W}}^k}(x^{k+1}, y^{k+1}) \tag{4.12}$$

such that

$$(g_x^k, g_y^k) + \nu_k(\mathcal{M}_{x,k}(x^{k+1} - \hat{x}^k), \mathbf{M}_{y,k}(y^{k+1} - \hat{y}^k)) = \mathbf{0}. \tag{4.13}$$

Comparing (4.13) with the expressions of (x^{k+1}, y^{k+1}), we obtain

$$(g_x^k, g_y^k) = (\mathcal{Q}(x^{k+1} - w^{k+1}), b - \mathcal{A}(w^{k+1})) \in \partial f_{\widehat{\mathcal{W}}^k}(x^{k+1}, y^{k+1}). \tag{4.14}$$

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

$$\bar{f}^k(x, y) = f_{w^{k+1}}(x^{k+1}, y^{k+1}) + \langle g_x^k, x - x^{k+1}\rangle + \langle g_y^k, y - y^{k+1}\rangle. \tag{4.15}$$

In addition,

$$(g_x^k, g_y^k) \in \partial_{\varepsilon_k} f(\hat{x}^k, \hat{y}^k), \tag{4.16}$$

where

$$0 \le \varepsilon_k := f(\hat{x}^k, \hat{y}^k) - \bar{f}^k(\hat{x}^k, \hat{y}^k). \tag{4.17}$$

To see this, note that in view of (4.14) and Proposition 4.2, for any $(x, y) \in \mathbb{V} \times \mathbb{R}^m$,

$$f_{\widehat{\mathcal{W}}^k}(x, y) \ge f_{\widehat{\mathcal{W}}^k}(x^{k+1}, y^{k+1}) + \langle g_x^k, x - x^{k+1}\rangle + \langle g_y^k, y - y^{k+1}\rangle = \bar{f}^k(x, y),$$

so that

$$\varepsilon_k \ge f_{\widehat{\mathcal{W}}^k}(\hat{x}^k, \hat{y}^k) - \bar{f}^k(\hat{x}^k, \hat{y}^k) \ge 0.$$

Hence

$$\begin{aligned} & f(\hat{x}^k, \hat{y}^k) + \langle x - \hat{x}^k, \mathcal{Q}(x^{k+1} - w^{k+1})\rangle + \langle y - \hat{y}^k, b - \mathcal{A}(w^{k+1})\rangle - \varepsilon_k \\ =& f_{\widehat{\mathcal{W}}^k}(x^{k+1}, y^{k+1}) + \langle x - x^{k+1}, \mathcal{Q}(x^{k+1} - w^{k+1})\rangle + \langle y - y^{k+1}, b - \mathcal{A}(w^{k+1})\rangle \\ \overset{(4.12)}{\le}& f_{\widehat{\mathcal{W}}^k}(x, y) \le f(x, y). \end{aligned}$$

The approximate subgradient relation in (4.16) can certainly be employed to derive an optimality estimate. In doing so, owing to (4.17), one needs to compute $f(\hat{x}^k, \hat{y}^k)$ 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 4.1. *Suppose that* $(\lambda_{ap}^{k+1}, \tilde{c}^{k+1}, \epsilon^{k+1})$ *and* $(\hat{\lambda}_{ap}^{k}, \hat{c}^{k}, \hat{\epsilon}^{k})$ *are Ritz-type triplets for* $\lambda_1(-\hat{s} - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(x^{k+1}))$ *and* $\lambda_1(-\hat{s} - \mathcal{A}^*(\hat{y}^k) - \mathcal{Q}(\hat{x}^k))$, *respectively. Then we have*

$$(g_x^k, g_y^k) \in \partial_{\eta^k} f(x^{k+1}, y^{k+1}) \quad \textit{and} \quad (g_x^k, g_y^k) \in \partial_{\tau^k} f(\hat{x}^k, \hat{y}^k) \tag{4.18}$$

where

$$\eta^k = f_u(x^{k+1}, y^{k+1}) - f_{w^{k+1}}(x^{k+1}, y^{k+1}), \tag{4.19}$$

$$\tau^k = f_l(\hat{x}^k, \hat{y}^k) + \alpha\hat{\epsilon}^k - \bar{f}^k(\hat{x}^k, \hat{y}^k) \geq 0. \tag{4.20}$$

Proof. In view of the property of the Ritz-type triplet for $\lambda_1(-\hat{s} - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(x^{k+1}))$ described in Proposition 3.4, for any $(x, y) \in \mathbb{V} \times \mathbb{R}^m$, it holds

$$\begin{aligned}
f(x, y) \geq & f_{\widehat{\mathcal{W}}^k}(x, y) \overset{(4.12)}{\geq} f_{\widehat{\mathcal{W}}^k}(x^{k+1}, y^{k+1}) + \langle g_x^k, x - x^{k+1}\rangle + \langle g_y^k, y - y^{k+1}\rangle \\
\geq & \alpha\lambda_1(-\hat{s} - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(x^{k+1})) + \langle g_x^k, x - x^{k+1}\rangle + \langle g_y^k, y - y^{k+1}\rangle \\
& - (\alpha\lambda_{ap}^{k+1} + \alpha\epsilon^{k+1} - f_{\widehat{\mathcal{W}}^k}(x^{k+1}, y^{k+1})) \\
\overset{(4.9)}{=} & f(x^{k+1}, y^{k+1}) + \langle g_x^k, x - x^{k+1}\rangle + \langle g_y^k, y - y^{k+1}\rangle - (\alpha\lambda_{ap}^{k+1} + \alpha\epsilon^{k+1} \\
& - \langle w^{k+1}, -\hat{s} - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(x^{k+1})\rangle) \\
\overset{(3.13),(3.14)}{=} & f(x^{k+1}, y^{k+1}) + \langle g_x^k, x - x^{k+1}\rangle + \langle g_y^k, y - y^{k+1}\rangle - f_u(x^{k+1}, y^{k+1}) + f_{w^{k+1}}(x^{k+1}, y^{k+1}).
\end{aligned}$$

Since $w^{k+1} \in \alpha\mathcal{W}$, it follows

$$\langle w^{k+1}, -\hat{s} - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(x^{k+1})\rangle \leq \alpha\lambda_1(-\hat{s} - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(x^{k+1})),$$

whence $\eta^k \geq 0$ and $(g_x^k, g_y^k) \in \partial_{\eta^k} f(x^{k+1}, y^{k+1})$.

Putting together (4.17) and the property of the Ritz-type triplet for the largest eigenvalue $\lambda_1(-\hat{s} - \mathcal{A}^*(\hat{y}^k) - \mathcal{Q}(\hat{x}^k))$ described in Proposition 3.4, we obtain $\tau^k \geq \varepsilon_k \geq 0$, whence $(g_x^k, g_y^k) \in \partial_{\tau^k} f(\hat{x}^k, \hat{y}^k)$ is satisfied by (4.16). $\square$

Set

$$d^k := (d_x^k, d_y^k) := (x^{k+1} - \hat{x}^k, y^{k+1} - \hat{y}^k), \ l_d^k = \|d_x^k\|^2_{\mathcal{M}_{x,k}} + \|d_y^k\|^2_{\mathbf{M}_{y,k}}. \tag{4.21}$$

In accordance with (4.13), we have, for $i = 1, 2$,

$$\begin{aligned}
&\langle g_x^k, x - x^i\rangle + \langle g_y^k, y - y^i\rangle = \nu_k(\langle \mathcal{M}_{x,k}(\hat{x}^k - x^{k+1}), x - x^i\rangle + \langle \mathbf{M}_{y,k}(\hat{y}^k - y^{k+1}), y - y^i\rangle)\\
=&\nu_k(\langle \mathcal{M}_{x,k}^{1/2}(\hat{x}^k - x^{k+1}), \mathcal{M}_{x,k}^{1/2}(x - x^i)\rangle + \langle \mathbf{M}_{y,k}^{1/2}(\hat{y}^k - y^{k+1}), \mathbf{M}_{y,k}^{1/2}(y - y^i)\rangle)\\
\leq&\|\nu_k \mathcal{M}_{x,k}^{1/2}(d_x^k)\|_F \|\mathcal{M}_{x,k}^{1/2}\|_2 \|x - x^i\|_F + \|\nu_k \mathbf{M}_{y,k}^{1/2} d_y^k\|_2 \|\mathbf{M}_{y,k}^{1/2}\|_2 \|y - y^i\|_2\\
\leq&\varrho_k(\|\nu_k \mathcal{M}_{x,k}^{1/2}(d_x^k)\|_F^2 + \|\nu_k \mathbf{M}_{y,k}^{1/2} d_y^k\|_2^2)^{1/2}(\|x - x^i\|_F^2 + \|y - y^i\|_2^2)^{1/2}\\
=&\varrho_k(\|\mathcal{M}_{x,k}^{-1/2}(g_x^k)\|_F^2 + \|\mathbf{M}_{y,k}^{-1/2} g_y^k\|_2^2)^{1/2}(\|x - x^i\|_F^2 + \|y - y^i\|_2^2)^{1/2}\\
=&\varrho_k \nu_k \sqrt{l_d^k}(\|x - x^i\|_F^2 + \|y - y^i\|_2^2)^{1/2},
\end{aligned}$$

where

$$\varrho_k = \max\{\|\mathcal{M}_{x,k}^{1/2}\|_2, \|\mathbf{M}_{y,k}^{1/2}\|_2\} > 0, \tag{4.22}$$

$(x^1, y^1) = (\hat{x}^k, \hat{y}^k)$ 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 (4.18) yield the optimality estimate

$$f(\hat{x}^k, \hat{y}^k) \leq f(x, y) + \varrho_k \nu_k \sqrt{l_d^k}(\|x - \hat{x}^k\|_F^2 + \|y - \hat{y}^k\|_2^2)^{1/2} + \tau^k \tag{4.23}$$

$$f(x^{k+1}, y^{k+1}) \leq f(x, y) + \varrho_k \nu_k \sqrt{l_d^k}(\|x - x^{k+1}\|_F^2 + \|y - y^{k+1}\|_2^2)^{1/2} + \eta^k \tag{4.24}$$

for all $(x, y) \in \mathbb{V} \times \mathbb{R}^m$.

The inequalities (4.23) and (4.24) say that the point $(\hat{x}^k, \hat{y}^k)$ or (x^{k+1}, y^{k+1}) is optimal if the optimality measure

$$\varpi^k = \max\left\{\varrho_k \nu_k \sqrt{l_d^k}, \ \min\{\eta^k, \tau^k\}\right\}, \tag{4.25}$$

is zero, a result 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 4.3. *For $w \in \widehat{\mathcal{W}}^k$, let $x^k_{min}(w)$ and $y^k_{min}(w)$ be defined by (4.10) and (4.11), respectively, it holds*

$$\begin{aligned} L^k(x,y,w) =& L^k(x^k_{min}(w), y^k_{min}(w), w) + \\ & \frac{1}{2}\|x - x^k_{min}(w)\|^2_{\mathcal{T}_k} + \frac{\nu_k}{2}\|y - y^k_{min}(w)\|^2_{\mathbf{M}_{y,k}}. \end{aligned} \tag{4.26}$$

Moreover, if (x^{k+1}, y^{k+1}) satisfies (4.7) and w^{k+1} is an optimal solution of (SQSCP), then

$$L^k(x^{k+1}, y^{k+1}, w^{k+1}) \le f(\hat{x}^k, \hat{y}^k), \tag{4.27}$$

and if, in addition, $\varpi^k = 0$ then $f(\hat{x}^k, \hat{y}^k) = f_{w^{k+1}}(x^{k+1}, y^{k+1})$ and $(\hat{x}^k, \hat{y}^k)$ is optimal.

Proof. By (4.5), together with (4.10) and (4.11), it follows

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

From Proposition 4.2, this yields

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

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

Since both $\mathcal{M}_{x,k}$ and $\mathbf{M}_{y,k}$ are invertible, by $l_d^k = 0$, we get $x^{k+1} = \hat{x}^k$ and $y^{k+1} = \hat{y}^k$. Furthermore, in combination with (4.13), we have $(g_x^k, g_y^k) = (\mathbf{0}, \mathbf{0})$. Together with (4.18), it then follows $(\mathbf{0}, \mathbf{0}) \in \partial f(x^{k+1}, y^{k+1}) = \partial f(\hat{x}^k, \hat{y}^k)$ because at least one of η^k and τ^k is equal to 0. Hence $(\hat{x}^k, \hat{y}^k)$ is optimal.

If $\eta^k = 0$, using (4.19) and Proposition 3.4, we have

$$\alpha\lambda_1(-\hat{s} - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(x^{k+1})) \le \langle w^{k+1}, -\hat{s} - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(x^{k+1})\rangle.$$

Consequently, $f_{w^{k+1}}(x^{k+1}, y^{k+1}) = f(x^{k+1}, y^{k+1}) = f(\hat{x}^k, \hat{y}^k)$ since $x^{k+1} = \hat{x}^k$ and $y^{k+1} = \hat{y}^k$.

If $\tau^k = 0$, by virtue of (4.15), (4.20) and Proposition 3.4, we obtain

$$f(\hat{x}^k, \hat{y}^k) \le f_{w^{k+1}}(x^{k+1}, y^{k+1}) \le L^k(x^{k+1}, y^{k+1}, w^{k+1}),$$

hence $f(\hat{x}^k, \hat{y}^k) = f_{w^{k+1}}(x^{k+1}, y^{k+1})$ by (4.27). □

4.2.3 Updating the model $\widehat{\mathcal{W}}^k$

In this subsection, we exploit aggregation to keep the rank r_k of $\mathbb{V}^{(r_k)}$ small. The technique is to compress the indispensable information into a single aggregate subgradient.

Suppose that we are given an approximate subgradient of $\lambda_1(-\hat{s} - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(x^{k+1}))$ at iteration k, via the scheme of the proximal bundle method, we could wish that this new approximate subgradient information can be added into the updated model $\widehat{\mathcal{W}}^{k+1}$. To this end, we need the following ingredients. First, we shall present an orthonormalization process, see Algorithm 4.3, to create a system of mutually orthogonal primitive idempotents, which in turn further generate the underlying subalgebra of $\widehat{\mathcal{W}}^{k+1}$.

Let $(v^\star, \xi^\star)$ be the optimal solution for the subproblem. Apply the spectral decomposition theorem in the subalgebra $\mathbb{V}_k^{(r_k)}$ to decompose $v^\star$ into

$$v^\star = \sum_{i=1}^{r_k} \lambda_i^k(v^\star, \mathbb{V}_k^{(r_k)}) c_i^k,$$

where $\{c_1^k, \cdots, c_{r_k}^k\}$ is a Jordan frame in $\mathbb{V}_k^{(r_k)}$.

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

$$\begin{aligned} &\{c_{1,k+1}, \cdots, c_{r_{k+1},k+1}\} := orth\{c_1^k, \cdots, c_{l_k}^k, \tilde{c}_{k+1}\}, \qquad (4.28)\\ &e_{r_{k+1}}^{k+1} = c_{1,k+1} + \cdots + c_{r_{k+1},k+1},\\ &\mathbb{V}_{k+1}^{(r_{k+1})} = \left\{x \in \mathbb{V} \mid x \circ e_{r_{k+1}}^{k+1} = x\right\},\\ &\mathbb{K}^{(r_{k+1})} = \mathbb{V}_{k+1}^{(r_{k+1})} \cap \mathbb{K}, \end{aligned}$$

where $1 \leq r_{k+1} \leq r$, and $orth\{a_1, a_2, \cdots, a_m\}$ is the orthonormalization of a finite set of elements $\{a_1, a_2, \cdots, a_m\} \subseteq \mathbb{V}$. It is significant to note that here orthogonality is with respect to the Jordan product. In what follows, we will see that how to achieve (4.28).

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

Proposition 4.4. *For every nonzero $u \in \mathbb{V}$ and a primitive idempotent $c \in \mathbb{V}$, there exist a primitive idempotent $a \in \mathbb{V}$ such that*

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

Moreover, if u is a scalar multiple of a primitive idempotent, then $P_u(c) = \mathbf{0}$ if and only if $\langle u, c\rangle = 0$.

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

Now we suppose that $P_u(c) \neq \mathbf{0}$. By a characterization of eigenvalues of $P_{P_u(c)}$ given by (iii) of Lemma 12 in [90] and (iii) of Corollary 2.1, it suffices to show that $P_{P_u(c)}$ has exactly one nonzero eigenvalue.

In view of (ii) in Proposition 2.2, we have

$$P_{P_u(c)} = P_u P_c P_u.$$

Denote $dim(\mathbb{V})$ by n. By means of (iii) of [90, Lemma 12], the eigenvalues of P_c are 1 and 0. Since $\mathbb{V}(c,1)$ is the eigenspace of P_c corresponding to the eigenvalue 1, together with (ii) of Corollary 2.1, it deduces that P_c has eigenvalues 1 with multiplicity one and 0 with multiplicity $n-1$. Thus, it is possible to expand the system $\{c\}$ to a canonical orthonormal (w.r.t. the trace inner product) basis $\{c, u_1, \cdots, u_{n-1}\}$ of $\mathbb{V}$, such that the matrix representation of P_c with respect to this basis is $Diag(1, 0, \cdots, 0)$. Suppose that the matrix representation of P_u with respect to this basis is A, then the matrix representation of $P_{P_u(c)}$ with respect to this basis is $ADiag(1, 0, \cdots, 0)A$. Since $u \neq \mathbf{0}$, we can assume without loss of generality that the first column of A is nonzero. By virtue of A being symmetric, the rank of $ADiag(1, 0, \cdots, 0)A$ is one. Therefore $ADiag(1, 0, \cdots, 0)A$ has exactly one nonzero eigenvalue. The second statement follows from Proposition 2.5. The proof is complete. □

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

Algorithm 4.2. Input : *a set of linearly independent primitive idempotents* $\{a_1, \cdots, a_k\}$, $2 \leq k \leq r$, *where* $a_1, \cdots, a_{k-1}$ *are mutually orthogonal primitive idempotents.*

Output: *a set of mutually orthogonal primitive idempotents*

$$\{c_1, \cdots, c_l\} := orth\{a_1, \cdots, a_k\},\ l \leq k,$$

which is a Jordan frame of the subalgebra generated by $\{a_1, \cdots, a_k\}$.

Perform the following steps.

1. *Set* $c_i = a_i,\ i = 1, \cdots, k-1$.

2. *Set* $c' = e - c_1 - \cdots - c_{k-1}$. *If* $\langle c', a_k \rangle = 0$, *then* $l = k-1$, *stop. Otherwise, set*

$$c_k = \frac{1}{1 - \sum_{i=1}^{k-1} \langle c_i, a_k \rangle} \left[a_k + \sum_{i=1}^{k-1} \langle c_i, a_k \rangle c_i + 4 \sum_{1 \leq i < j \leq k-1} L_{c_i}(L_{c_j}(a_k)) - 2 \sum_{i=1}^{k-1} L_{c_i}(a_k) \right]. \tag{4.29}$$

We next present a process to orthonormalize a finite set of linearly independent primitive idempotents. The process takes k linearly independent primitive idempotents to an orthonormal set that spans a subalgebra of $\mathbb{V}$, whose rank is at most k.

Algorithm 4.3. Input : *a set of linearly independent primitive idempotents* $\{a_1, \cdots, a_k\}$, $2 \le k \le r$.

Output: *a set of mutually orthogonal primitive idempotents*

$$\{c_1, \cdots, c_l\} := orth\{a_1, \cdots, a_k\}, \; l \le k,$$

which is a Jordan frame of the subalgebra generated by $\{a_1, \cdots, a_k\}$.

Set $c_1 := a_1$ *and* $S_1 := \{c_1\}$.

For $i = 2, \cdots, k,$ *do*

$S_i := orth\{S_{i-1}, a_i\}$, *generated by Algorithm 4.2. Denote the cardinality of* S_i *by* $\kappa(i)$, *and we list the elements in* S_i *as* $\{c_1, \cdots, c_{\kappa(i)}\}$.

end

The following result tells us that idempotents generated by Algorithm 4.3 are well defined, which, in turn, establishes that the new approximate subgradient $\tilde{c}_{l+1}$ has the desired property

$$\tilde{c}_{k+1} \in \mathbb{V}_{k+1}^{(r_k+1)}. \tag{4.30}$$

Theorem 4.1. *Suppose that* $a_1, \cdots, a_k$ $(2 \le k \le r)$ *are linearly independent primitive idempotents, then*

$$\{c_1, \cdots, c_l\} := orth\{a_1, \cdots, a_k\}, \; l \le k,$$

generated by Algorithm 4.3, is well defined, and for $i = 1, \cdots, k$,

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

Proof. We first show that Algorithm 4.2 works, which, in turn, shows that Algorithm 4.3 is well defined. For a set of mutually orthogonal primitive idempotents $\{c_1, \cdots, c_l\}$, in light of (ii) of Proposition 2.9, we have

$$\{x \in \mathbb{V} \mid x \circ (c_1 + \cdots + c_i) = x\} \subseteq \{x \in \mathbb{V} \mid x \circ (c_1 + \cdots + c_i + c_{i+1}) = x\}$$

for $1 \leq i \leq l-1$. Together with the design of Algorithm 4.2, it then suffices to prove that

$$\{c_1, \cdots, c_{\kappa(i)}\} := orth\{c_1, \cdots, c_{\kappa(i-1)}, a_i\}$$

is well defined for every fixed i in $\{2, \cdots, k\}$, where $c_1, \cdots, c_{\kappa(i-1)}$ are mutually orthogonal primitive idempotents, and

$$a_i \in \mathbb{V}^{(\kappa(i))} = \{x \in \mathbb{V} \mid x \circ (c_1 + \cdots + c_{\kappa(i)}) = x\}, \ \kappa(i) \leq i.$$

Thus without loss of generality we may assume $i = k$, $\kappa(i) = l$ and $\kappa(i-1) = k-1$.

Set $c' = e - c_1 - \cdots - c_{k-1}$. By virtue of Proposition 4.4, there exists a primitive idempotent $c_k \in \mathbb{V}$ such that $P_{c'}(a_k) = \langle c', a_k \rangle c_k =: \gamma c_k$. We see that $P_{c'}(a_k) = \mathbf{0}$ if and only if $\langle c', a_k \rangle = 1 - \sum_{i=1}^{k-1} \langle c_i, a_k \rangle = 0$ by Proposition 4.4. In this case c_k can be any primitive idempotent which is orthogonal to $c_1, \cdots, c_{k-1}$, and we set $l = k-1$. Otherwise,

$$\langle c_k, c_i \rangle = \frac{1}{\gamma} \langle a_k, P_{c'}(c_i) \rangle = 0, \ i = 1, \cdots, k-1,$$

since $\langle c', c_i \rangle = 0$ and then $P_{c'}(c_i) = \mathbf{0}$ by Proposition 4.4. We set $l = k$. By Proposition 2.5, it yields $c_k \circ c_i = \mathbf{0}, \ i = 1, \cdots, k-1$. Hence, it is possible to complete the system $\{c_1, \cdots, c_k\}$ to a Jordan frame $\{c_1, \cdots, c_k, \cdots, c_r\}$ of $\mathbb{V}$, and

$$c' = c_k + \cdots + c_r. \tag{4.31}$$

Using the Peirce decomposition Theorem 2.4, we obtain

$$a_k = \sum_{i=1}^{k} x_i c_i + \sum_{1 \leq i < j \leq k} x_{ij} + \sum_{i=k+1}^{r} x_i c_i + \sum_{k+1 \leq i < j \leq r} x_{ij} + \sum_{1 \leq i \leq k < j \leq r} x_{ij}. \tag{4.32}$$

Applying $P_{c'}$ on the both sides of (4.32), together with (4.31) and (v) in Proposition

2.2 , we then obtain

$$\gamma c_k = \sum_{i=k}^{r} x_i c_i + \sum_{k \le i < j \le r} x_{ij}.$$

Taking the inner product with c_i on the both sides of the above equation with $i = k, \cdots, r$, we have

$$x_k = \gamma \text{ and } x_i = 0, i = k+1, \cdots, r, \tag{4.33}$$

whence $\sum_{k \le i < j \le r} x_{ij} = \mathbf{0}$, which follows $\langle a_k, c_i \rangle = 0$ by taking the inner product with c_i on the both sides of (4.32), $i = k+1, \cdots, r$. Therefore, $a_k \circ c_i = \mathbf{0},\ i = k+1, \cdots, r$ by Proposition 2.5. Thus, we have

$$a_k \in \begin{cases} \{x \in \mathbb{V} \mid x \circ (c_1 + \cdots + c_{k-1}) = x\} & \text{if } \gamma = 0, \\ \{x \in \mathbb{V} \mid x \circ (c_1 + \cdots + c_k) = x\} & \text{if } \gamma \ne 0. \end{cases}$$

Furthermore, taking the product $\circ$ with c_i on the both sides of (4.32) for $i = k+1, \cdots, r$, and then summing them up on both sides, we obtain

$$\sum_{k+1 \le i < j \le r} x_{ij} + \frac{1}{2} \sum_{1 \le i \le k < j \le r} x_{ij} = \mathbf{0}.$$

On the other hand, $\sum_{k \le i < j \le r} x_{ij} = \mathbf{0}$ and $\sum_{j=k+1}^{r} x_{kj} = 2c_k \circ (\sum_{k \le i < j \le r} x_{ij}) = c_k \circ \mathbf{0} = \mathbf{0}$ imply

$$\sum_{k+1 \le i < j \le r} x_{ij} = \mathbf{0}, \tag{4.34}$$

thus

$$\sum_{1 \le i \le k < j \le r} x_{ij} = \mathbf{0}. \tag{4.35}$$

We shall prove the equation (4.29) is valid for $\gamma \ne 0$.

In combination with (4.33), (4.34) and (4.35), we can further rewrite a_k to

$$a_k = \sum_{i=1}^{k-1} x_i c_i + \sum_{1 \le i < j \le k-1} x_{ij} + \gamma c_k + \sum_{i=1}^{k-1} x_{ik},$$

where $x_i = \langle c_i, a_k \rangle,\ i = 1, \cdots, k-1$. Meanwhile, $1 = tr(a_k) = \sum_{i=1}^{k-1} x_i + \gamma$ implies that

$$\gamma = 1 - \sum_{i=1}^{k-1} \langle c_i, a_k \rangle.$$

Since $(\sum_{i=1}^{k-1} c_i) \circ (\sum_{1 \le i<j \le k-1} x_{ij}) = \sum_{1 \le i<j \le k-1} x_{ij}$ by (i) and (ii) of Proposition 2.9, we have

$$\sum_{i=1}^{k-1} x_{ik} = 2\Big[\sum_{i=1}^{k-1} c_i \circ a_k - \sum_{i=1}^{k-1} x_i c_i - \sum_{1 \le i<j \le k-1} x_{ij}\Big],$$

which proves (4.29). □

We are now on a position to provide an orthonormalization procedure for (4.28).

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

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

Obviously, (4.30) holds.

Case 2. $c_1^k, \cdots, c_{l_k}^k, \tilde{c}_{k+1}$ are linearly independent. In this case $\{c_{1,k+1}, \cdots, c_{r_{k+1},k+1}\}$ can be generated by Algorithm 4.2, where $r_{k+1} \le l_k + 1$. Again, (4.30) holds.

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

$$\bar{w}_{k+1} = \frac{\sum_{i=l_k+1}^{r_k} \lambda_i^k c_i^k + \xi^\star \bar{w}_k}{\sum_{i=l_k+1}^{r_k} \lambda_i^k + \xi^\star}. \tag{4.36}$$

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

Clearly, $\bar{v} = \sum_{i=1}^{l_k} \lambda_i^k c_{i,k+1} \in \mathbb{K}^{(r_{k+1})}$ and (4.28) implies $\tilde{c}_{k+1} \in \mathbb{V}_{k+1}^{(r_{k+1})}$ by Theorem 4.1. Therefore

$$\begin{aligned} w_{ap}^{k+1} &:= \alpha \tilde{c}_{k+1} \in \widehat{\mathcal{W}}^{k+1}, \\ w^{k+1} &= v^\star + \xi^\star \bar{w}_k = \bar{v} + (\sum_{i=l_k+1}^{r_k} \lambda_i^k + \xi^\star) \bar{w}_{k+1} \in \widehat{\mathcal{W}}^{k+1}. \end{aligned}$$

We can now summarize the above analysis as follows.

Proposition 4.5. *For $w_{ap}^{k+1} = \alpha\tilde{c}_{k+1}$, update formulas (4.28) and (4.36) ensure that both w_{ap}^{k+1} and w^{k+1} belong to $\widehat{\mathcal{W}}^{k+1}$ of (4.3).*

We will see from the convergence analysis that the property of w_{ap}^{k+1} and w^{k+1} described in the above proposition is all that is needed to guarantee convergence.

4.2.4 Updating the weight ν_k

The choice of weight ν_k is somewhat of an art, which impacts significantly both global convergence in theory and efficiency in practice. For instance, as indicated in [40], a suitable adjustment of ν_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, [11, 49, 91]. The updating rule here is the same as [49], we outline the main results here.

First we detect whether ν_k is too large.

Let $d := (d_x, d_y) := (x - \hat{x}^k, y - \hat{y}^k)$. It will be convenient to have an alternate description of $f_w(x, y)$, which is in terms of d. With the help of (4.4), it yields

$$\begin{aligned} f_w(x,y) &= f_w(\hat{x}^k, \hat{y}^k) + \langle b - \mathcal{A}(w), d_y\rangle + \langle \mathcal{Q}(\hat{x}^k - w), d_x\rangle + \frac{1}{2}\langle d_x, \mathcal{Q}(d_x)\rangle \\ &=: f_w^{dr}((\hat{x}^k, \hat{y}^k), d). \end{aligned}$$

Correspondingly, $f_{\widehat{\mathcal{W}}^k}^{dr}((\hat{x}^k, \hat{y}^k), d) = \max_{w\in\widehat{\mathcal{W}}^k} f_w^{dr}((\hat{x}^k, \hat{y}^k), d)$, thus

$$f_{\widehat{\mathcal{W}}^k}^{dr}((\hat{x}^k, \hat{y}^k), \mathbf{0}) = f_{\widehat{\mathcal{W}}^k}(\hat{x}^k, \hat{y}^k) \le f(\hat{x}^k, \hat{y}^k).$$

By virtue of the convexity of $f_{\widehat{\mathcal{W}}^k}^{dr}((\hat{x}^k, \hat{y}^k), d)$ as a function of d and $f_l(\hat{x}^k, \hat{y}^k) \le f(\hat{x}^k, \hat{y}^k)$, in combination with $d^k = (x^{k+1} - \hat{x}^k, y^{k+1} - \hat{y}^k)$, we obtain, for $\kappa \in [0, 1]$,

$$\begin{aligned} & f_{\widehat{\mathcal{W}}^k}^{dr}((\hat{x}^k, \hat{y}^k), \kappa d^k) \\ \le& (1-\kappa) f_{\widehat{\mathcal{W}}^k}^{dr}((\hat{x}^k, \hat{y}^k), \mathbf{0}) + \kappa f_{\widehat{\mathcal{W}}^k}^{dr}((\hat{x}^k, \hat{y}^k), d^k) \\ \overset{(4.9)}{=}& (1-\kappa) f_{\widehat{\mathcal{W}}^k}(\hat{x}^k, \hat{y}^k) + \kappa f_{w^{k+1}}(x^{k+1}, y^{k+1}) \\ \le& (1-\kappa) f(\hat{x}^k, \hat{y}^k) + \kappa f_{w^{k+1}}(x^{k+1}, y^{k+1}) \\ \le& f(\hat{x}^k, \hat{y}^k) - \kappa\vartheta^k, \end{aligned}$$

with

$$\vartheta^k = f_l(\hat{x}^k, \hat{y}^k) - f_{w^{k+1}}(x^{k+1}, y^{k+1}). \tag{4.37}$$

Therefore ϑ^k estimates the descent obtained from our model and will subsequently serve as *predicted descent* of f. We emphasize that ϑ^k may be nonpositive and $f_l(\hat{x}^k, \hat{y}^k) - f_u(x^{k+1}, y^{k+1}) \geq \kappa\vartheta^k$ can guarantee $\vartheta^k > 0$ for $\kappa \in (0, 1)$. Alternatively, we can rewrite the predicted descent ϑ^k in terms of τ^k, $\hat{\epsilon}^k$ and ν_k, that is,

$$\begin{aligned}
\vartheta^k &\overset{(4.37)}{=} f_l(\hat{x}^k, \hat{y}^k) + \alpha\hat{\epsilon}^k - \bar{f}^k(\hat{x}^k, \hat{y}^k) - \alpha\hat{\epsilon}^k + \bar{f}^k(\hat{x}^k, \hat{y}^k) - f_{w^{k+1}}(x^{k+1}, y^{k+1}) \\
&\overset{(4.20),(4.15)}{=} \tau^k - \alpha\hat{\epsilon}^k + \langle g_x^k, \hat{x}^k - x^{k+1}\rangle + \langle g_y^k, \hat{y}^k - y^{k+1}\rangle \\
&\overset{(4.13),(4.21)}{=} \tau^k - \alpha\hat{\epsilon}^k + \nu_k l_d^k.
\end{aligned} \tag{4.38}$$

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

The weight ν_k 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

$$f_l(\hat{x}^k, \hat{y}^k) - f_u(x^{k+1}, y^{k+1}) \geq m_R\vartheta^k \tag{4.39}$$

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

$$\nu_{k+1}^{int} = 2\nu_k\left[1 - \frac{f_l(\hat{x}^k, \hat{y}^k) - f_u(x^{k+1}, y^{k+1})}{\vartheta^k}\right], \tag{4.40}$$

we have $\frac{\nu_{k+1}^{int}}{\nu_k} \leq 2(1 - m_R) < 1$ if (4.39) is satisfied. In this case, we set

$$\nu_{k+1} = \max\left\{\nu_{k+1}^{int}, \frac{\nu_k}{10}, \nu_{min}\right\}, \tag{4.41}$$

where ν_{min} is a small positive constant.

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

We define the error of new approximation by

$$e_v^k = f_u(\hat{x}^k, \hat{y}^k) - f_{w_{ap}^{k+1}}(\hat{x}^k, \hat{y}^k), \tag{4.42}$$

and the variation

$$(var)_k = f(\hat{x}^k, \hat{y}^k) - \min \left\{ f(x, y) \mid (x, y) \in \mathscr{B}_{\mathbf{M}}(\hat{x}^k, \hat{y}^k, l_d^k) \right\},$$

where

$$\mathscr{B}_{\mathbf{M}}(\hat{x}^k, \hat{y}^k, l_d^k) = \left\{ (x, y) \in \mathbb{V} \times \mathbb{R} \;\middle|\; \|x - \hat{x}^k\|^2_{\mathcal{M}_{x,k}} + \|y - \hat{y}^k\|^2_{\mathbf{M}_{y,k}} \leq l_d^k \right\}.$$

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

$$\begin{aligned}
0 \leq (var)_k \leq & f(\hat{x}^k, \hat{y}^k) - \min \left\{ f_{\widehat{\mathcal{W}}^k}(x, y) \mid (x, y) \in \mathscr{B}_{\mathbf{M}}(\hat{x}^k, \hat{y}^k, l_d^k) \right\} \\
\leq & f(\hat{x}^k, \hat{y}^k) - \min \left\{ f_{w^{k+1}}(x, y) \mid (x, y) \in \mathscr{B}_{\mathbf{M}}(\hat{x}^k, \hat{y}^k, l_d^k) \right\} \\
= & f(\hat{x}^k, \hat{y}^k) - f_{w^{k+1}}(x^{k+1}, y^{k+1}) \\
\leq & f_l(\hat{x}^k, \hat{y}^k) - f_{w^{k+1}}(x^{k+1}, y^{k+1}) + \alpha \hat{\epsilon}^k = \vartheta^k + \alpha \hat{\epsilon}^k.
\end{aligned}$$

One may use the test

$$e_v^k > \vartheta^k + \alpha \hat{\epsilon}^k \tag{4.43}$$

to decide whether ν_k should be increased.

4.2.5 Description of the algorithm

In this subsection, we present an inexact spectral bundle method for the eigenvalue optimization problem (Eigform), combining ideas from [49], [70] and [38].

In what follows, the expression $[\lambda_{ap}, \tilde{c}, \epsilon] = \mathbf{eigest}(x, y, \delta)$ is the routine that produces a *Ritz-type triplet* $(\lambda_{ap}, \tilde{c}, \epsilon)$ for $\lambda_1(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x))$ such that $\alpha\epsilon \leq \delta$.

Algorithm 4.4. ***Input:*** *An initial point* $(x^0, y^0) \in \mathbb{V} \times \mathbb{R}^m$, *an* $\varepsilon \geq 0$ *for termination, an improvement parameter* $m_L \in (0, 0.5)$, *an initial subdifferential tolerance* δ^0, *an initial weight* $\nu_0 > \|\mathcal{Q}\|_2$, *a minimal weight* $\nu_{min} > \|\mathcal{Q}\|_2$, *set the variation estimate* $\theta_v^0 = \infty$.

1. ***(Initialization)*** *Set* $k = 0$, $\hat{x}^0 = x^0$, $\hat{y}^0 = y^0$, $s = 0$. *Call* $[\hat{\lambda}_{ap}^0, \hat{c}^0, \hat{\epsilon}^0] = \mathbf{eigest}(\hat{x}^0, \hat{y}^0, \delta^0)$. *Compute* (g_x^0, g_y^0), $\widehat{\mathcal{W}}^0$, *and* $f_l(\hat{x}^0, \hat{y}^0)$.

2. ***(Trial point finding)*** *Solve (SQSCP) to get* w^{k+1}*, compute* (x^{k+1}, y^{k+1}) *and* l_d^k *from (4.7) and (4.21), respectively.*

3. ***(Evaluation)*** *Call* $[\lambda_{ap}^{k+1}, \tilde{c}^{k+1}, \epsilon^{k+1}] = \mathbf{eigest}(x^{k+1}, y^{k+1}, \delta^k)$*. Compute* η^k*,* τ^k *and* ϖ^k *from (4.19), (4.20) and (4.25), respectively. Set* $w_{ap}^{k+1} = \alpha\tilde{c}^{k+1}$*.*

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

5. ***(Descent step)*** *If* $f_l(\hat{x}^k, \hat{y}^k) - f_l(x^{k+1}, y^{k+1}) \geq m_L\vartheta^k$*, then perform the following steps. Otherwise continue with step 6.*

   (i) Set $(\hat{x}^{k+1}, \hat{y}^{k+1}) = (x^{k+1}, y^{k+1}), \hat{\lambda}_{ap}^{k+1} = \lambda_{ap}^{k+1}, \hat{\epsilon}^{k+1} = \epsilon^{k+1}$*, and* $\hat{c}^{k+1} = \tilde{c}^{k+1}$*.*

   (ii) Select $\nu_{k+1} \in [\nu_{min}, \nu_k]$*(e.g. by (4.41)).*

   (iii) Set

$$\theta_v^{k+1} = \max\left\{\theta_v^k, 2(\vartheta^k + \alpha\hat{\epsilon}^k)\right\}. \tag{4.44}$$

   (iv) Pick $\delta^{k+1} \leq \delta^k$ *to ensure convergence.*

   (v) Choose the operator $\mathcal{M}_{x,k+1}$ *and the matrix* $\mathbf{M}_{y,k+1}$ *such that (1.3) holds.*

   (vi) Set $s = s + 1$ *and continue with step 7.*

6. ***(Null step)*** *Perform the following steps.*

   (i) Set $(\hat{x}^{k+1}, \hat{y}^{k+1}) = (\hat{x}^k, \hat{y}^k)$*.*

   (ii) If $\alpha\hat{\epsilon}^k > 10\delta^k$ *then call* $[\hat{\lambda}_{ap}^{k+1}, \hat{c}^{k+1}, \hat{\epsilon}^{k+1}] = \mathbf{eigest}(\hat{x}^k, \hat{y}^k, \delta^k)$*; otherwise set* $\hat{\lambda}_{ap}^{k+1} = \hat{\lambda}_{ap}^k, \hat{\epsilon}^{k+1} = \hat{\epsilon}^k$*, and* $\hat{c}^{k+1} = \hat{c}^k$*.*

   (iii) Set

$$\theta_v^{k+1} = \min\left\{\theta_v^k, \vartheta^k + \alpha\hat{\epsilon}^k\right\}. \tag{4.45}$$

   (iv) Either set $\nu_{k+1} = \nu_k$ *or choose* $\nu_{k+1} \in [\nu_k, 10\nu_k]$*(e.g.* $\nu_{k+1} = \min\left\{\nu_{k+1}^{int}, 10\nu_k\right\}$*) if*

$$e_v^k > \max\left\{\theta_v^k, 10(\vartheta^k + a\hat{\epsilon}^k)\right\}. \tag{4.46}$$

   (v) Set $\delta^{k+1} \leq \delta^k$ *to ensure convergence.*

   (vi) Choose the operator $\mathcal{M}_{x,k+1}$ *and the matrix* $\mathbf{M}_{y,k+1}$ *such that (1.4) is satisfied.*

7. ***(Updating*** $\widehat{\mathcal{W}}^k$***)*** *Choose a* $\widehat{\mathcal{W}}^{k+1} \supseteq \{w^{k+1}, w_{ap}^{k+1}\}$ *of the form (4.3).*

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

Remark 4.4. *(1). Set* $c_{10} = \tilde{c}^0$, $\bar{w}^0 = \alpha\tilde{c}^0$, *then compute* $\widehat{\mathcal{W}}^0$ *by the form (4.3).*

(2). To see the need for recomputing a Ritz-type triplet $[\hat{\lambda}_{ap}^{k+1}, \hat{c}^{k+1}, \hat{\epsilon}^{k+1}]$ *at step 6, suppose that, starting with iteration N, the implementation of Algorithm 4.4 is without the descent test. If the recomputation of the Ritz-type triplet was omitted, (ii) of step 6 would set* $(\hat{\lambda}_{ap}^{k+1}, \hat{c}^{k+1}, \hat{\epsilon}^{k+1}) = (\hat{\lambda}_{ap}^{k}, \hat{c}^{k}, \hat{\epsilon}^{k})$. *It would yield* $\alpha\hat{\epsilon}^k \leq \delta^{N-1}$ *for any* $k > N$, *which might prevent convergence. Furthermore, the recomputation can estimate the largest eigenvalue* $\lambda_1(-\hat{s} - \mathcal{A}^*(\hat{y}^k) - \mathcal{Q}(\hat{x}^k))$ *within a decreasing sequence of tolerances in the course of the number of null steps approaching infinity.*

(3). At iteration $k+1$, *we can choose* $\widehat{\mathcal{W}}^{k+1}$ *as follows. Decompose* $v^\star$ *into*

$$v^\star = \sum_{i=1}^{l_k} \lambda_i^k(v^\star, \mathbb{V}_k^{(r_k)}) c_i^k + \sum_{i=l_k+1}^{r_k} \lambda_i^k(v^\star, \mathbb{V}_k^{(r_k)}) c_i^k$$

where $\{c_1^k, \cdots, c_{r_k}^k\}$ *is a Jordan frame in* $\mathbb{V}_k^{(r_k)}$ *with* $l_k \leq r-1$. *Recall that* l_k *is the number of the "large" eigenvalues of* $v^\star$. *Compute* $\bar{w}_{k+1}$ *and* $\{c_{1,k+1}, \cdots, c_{r_{k+1},k+1}\}$ *by using (4.36) and Algorithm 4.2, respectively.*

(4). It has been pointed out in [11, Sect.10] that a sequence of null steps between two stability centers is just for the sake of the improvement of the model but possessing no new reliable information, 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 [58], and it is employed in [48]. At the same time, to the best of our knowledge, most of the literature about the bundle method exploits the special matrix* $\mathbf{M}_{y,k} = I$ *and excludes the variable* x, *see, for instance, [5, 8, 35, 38, 40, 49, 70, 91]. However, in order to ensure convergence, we should use variable metric at both descent steps and null steps in Algorithm 4.4.*

(5). Henceforth, we assume that the sequence $\{\|\mathcal{M}_{x,k}\|_2, \|\mathbf{M}_{y,k}\|_2\}_{k=1}^{\infty}$ *is bounded, any accumulation point of* $\{\mathcal{M}_{x,k}\}$ *is a self-adjoint positive definite linear operator, and any accumulation point of* $\{\mathbf{M}_{y,k}\}$ *is positive definite. More precisely, we make the following assumption:*

*(**A3**) If the number of null steps or descent steps in Algorithm 4.4 approaches infinity, then $\{\mathcal{M}_{x,k}\}$ converges to a self-adjoint positive definite linear operator and $\{\mathbf{M}_{y,k}\}$ converges to a positive definite matrix. Moreover, both (1.3) and (1.4) are satisfied.*

As a consequence, the sequence $\{\varrho_k\}_{k=1}^{\infty}$ is bounded.

4.3 Convergence analysis

In this section, we give a proof of the global convergence of the inexact spectral bundle method. Although the basic ideas of the proofs of convergence theorems are similar to those presented in [38], their details are more involved due to the existence of a quadratic term in the objective function of the problem.

We will prove the convergence of Algorithm 4.4 for $\varepsilon = 0$. If Algorithm 4.4 terminates after a finite number K of iterations, then by Proposition 4.3, $(\hat{x}^K, \hat{y}^K)$ is optimal. Thus the case of interest is when Algorithm 4.4 does not stop. The proof of convergence of the algorithm in this setting will be split into several lemmas and propositions. First we consider the case when only null step occurs after some iteration N.

In what follows, we assume that, starting with iteration N, the descent test fails in all subsequent iterations in Algorithm 4.4, i.e.

$$\hat{x} = \hat{x}^N = \hat{x}^{N+1} = \cdots, \ \hat{y} = \hat{y}^N = \hat{y}^{N+1} = \cdots.$$

For any $k > N$, we define

$$x^k = x^N_{min}(w^k) = \hat{x} + \mathcal{T}_{k-1}^{-1}(\mathcal{Q}(w^k - \hat{x})), \tag{4.47}$$

$$y^k = y^N_{min}(w^k) = \hat{y} + \frac{1}{\nu_{k-1}}\mathbf{M}_{y,k-1}^{-1}(\mathcal{A}(w^k) - b). \tag{4.48}$$

Lemma 4.1. *Let $\{(x^k, y^k)\}_{k=N}^{\infty}$ be specified by (4.47) and (4.48). If Assumption* **A3** *is satisfied, then*

$$\|x^{k+1} - x^k\|_F \to 0 \text{ and } \|y^{k+1} - y^k\|_2 \to 0 \text{ as } k \to \infty. \tag{4.49}$$

Proof. 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

$$\begin{aligned}
&L^k(x, y, w) \\
=&f_w(x, y) + \frac{\nu_k}{2}(\|x - \hat{x}\|^2_{\mathcal{M}_{x,k}} + \|y - \hat{y}\|^2_{\mathbf{M}_{y,k}}) \\
=&\langle -\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x), w\rangle + b^T y + \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle + \frac{\nu_k}{2}(\|x - \hat{x}\|^2_{\mathcal{M}_{x,k}} + \|y - \hat{y}\|^2_{\mathbf{M}_{y,k}}).
\end{aligned}$$

If there is no descent step from iteration N, then by step 6, $\nu_{k+1} \geq \nu_k$ for any $k \geq N$.

Using the definition of $\mathcal{T}_k$ and $\nu_k \geq \nu_{k-1}$, we have

$$\begin{aligned}
&L^k(x^{k+1}, y^{k+1}, w^k) \\
&\overset{(4.26)}{=} L^k(x^N_{min}(w^k), y^N_{min}(w^k), w^k) + \frac{1}{2}\|x^{k+1} - x^N_{min}(w^k)\|^2_{\mathcal{T}_k} + \frac{\nu_k}{2}\|y^{k+1} - y^N_{min}(w^k)\|^2_{\mathbf{M}_{y,k}} \\
&\overset{(4.47),(4.48)}{\geq} L^k(x^k, y^k, w^k) + \frac{\nu_k}{2}(\|x^{k+1} - x^k\|^2_{\mathcal{M}_{x,k}} + \|y^{k+1} - y^k\|^2_{\mathbf{M}_{y,k}}) \\
&\overset{(1.4)}{\geq} L^{k-1}(x^k, y^k, w^k) + \frac{\nu_{min}}{2}(\|x^{k+1} - x^k\|^2_{\mathcal{M}_{x,k}} + \|y^{k+1} - y^k\|^2_{\mathbf{M}_{y,k}}) \\
&\geq L^{k-1}(x^k, y^k, w^k).
\end{aligned}$$

Thus for any $k \geq N$,

$$\begin{aligned}
&L^{k-1}(x^k, y^k, w^k) \\
\leq&L^k(x^{k+1}, y^{k+1}, w^k) \\
\leq&L^k(x^{k+1}, y^{k+1}, w^{k+1}) \\
\leq&f(\hat{x}, \hat{y}),
\end{aligned} \tag{4.50}$$

where the middle inequality follows from (4.8) and $w^k \in \widehat{\mathcal{W}}^k$ by step 7, and the last inequality follows from (4.27). Therefore, there exists some $f_\star \in \mathbb{R}$ such that

$$L^{k-1}(x^k, y^k, w^k) \uparrow f_\star \leq f(\hat{x}, \hat{y}) \text{ as } k \to \infty, \tag{4.51}$$

whence

$$f_\star \le f_\star + \lim_{k\to\infty} \frac{\nu_{min}}{2}(\|x^{k+1}-x^k\|^2_{\mathcal{M}_{x,k}} + \|y^{k+1}-y^k\|^2_{\mathbf{M}_{y,k}}) \le f_\star.$$

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 (4.49) holds by Assumption **A3**. □

Lemma 4.2. *Let $\{(x^k, y^k)\}_{k=N}^\infty$ be specified by (4.47) and (4.48). If $\nu_k \uparrow \bar{\nu} \in (0,\infty)$ as $k\to\infty$ and*

$$f_{w^k}(x^k, y^k) \to f(\hat{x}, \hat{y}) \text{ as } k\to\infty, \tag{4.52}$$

then

$$(x^k, y^k) \to (\hat{x}, \hat{y}) \text{ as } k\to\infty, \tag{4.53}$$

$$b - \mathcal{A}(w^k) \to \mathbf{0} \quad \text{and} \quad \mathcal{Q}(x^k - w^k) \to \mathbf{0} \text{ as } k\to\infty. \tag{4.54}$$

Proof. Noting that $w^k \in \alpha\mathcal{W}$, we have

$$\begin{aligned}
&f(\hat{x}, \hat{y}) = \max_{w\in\alpha\mathcal{W}} f_w(\hat{x}, \hat{y}) \\
\ge& f_{w^k}(\hat{x}, \hat{y}) + \frac{\nu_{k-1}}{2}(\|\hat{x}-\hat{x}\|^2_{\mathcal{M}_{x,k-1}} + \|\hat{y}-\hat{y}\|^2_{\mathbf{M}_{y,k-1}}) \\
=& L^{k-1}(\hat{x}, \hat{y}, w^k) \overset{(4.26)}{=} L^{k-1}(x^k, y^k, w^k) + \frac{1}{2}\|\hat{x}-x^k\|^2_{\mathcal{T}_{k-1}} + \frac{\nu_{k-1}}{2}\|\hat{y}-y^k\|^2_{\mathbf{M}_{y,k-1}} \\
\ge& f_{w^k}(x^k, y^k) + \frac{\nu_{k-1}}{2}\|\hat{x}-x^k\|^2_{\mathbf{M}_{x,k-1}} + \frac{\nu_{k-1}}{2}\|\hat{y}-y^k\|^2_{\mathbf{M}_{y,k-1}}.
\end{aligned}$$

Then

$$\frac{\nu_{k-1}}{2}\|\hat{x}-x^k\|^2_{\mathcal{M}_{x,k-1}} + \frac{\nu_{k-1}}{2}\|\hat{y}-y^k\|^2_{\mathbf{M}_{y,k-1}} \le f(\hat{x}, \hat{y}) - f_{w^k}(x^k, y^k),$$

whence (4.53) holds by (4.52) and $\nu_k \uparrow \bar{\nu} > 0$. Therefore,

$$b - \mathcal{A}(w^k) = \nu_{k-1}\mathbf{M}_{y,k-1}(\hat{y}-y^k) \to \mathbf{0},$$

$$\mathcal{Q}(x^k - w^k) = \nu_{k-1}\mathcal{M}_{x,k-1}(\hat{x} - x^k) \to \mathbf{0},$$

as $k \to \infty$. □

Lemma 4.3. *Let* $\{(x^k, y^k)\}_{k=N}^{\infty}$ *be specified by (4.47) and (4.48). If Assumption* **A3** *and* $\lim_{k\to\infty} \delta^k = 0$ *are satisfied, then we have (4.52), (4.53) and (4.54).*

Proof. The proof of (4.52) hinges on the following observation:

$$\Delta_k := f(x^k, y^k) - f_{w^k}(x^k, y^k) \to 0 \text{ as } k \to \infty. \tag{4.55}$$

We will soon show it.

First of all, we see that (4.49) holds by Lemma 4.1.

Set

$$(g_{x\_ap}^{k-1}, g_{y\_ap}^{k-1}) := (\mathcal{Q}(x^k - w_{ap}^k), b - \mathcal{A}(w_{ap}^k)).$$

By step 3, Proposition 3.4 and (3.7), we have

$$(g_{x\_ap}^{k-1}, g_{y\_ap}^{k-1}) \in \partial_{\epsilon^k} f(x^k, y^k).$$

Define a linear function on $\mathbb{V} \times \mathbb{R}^m$ by

$$l^k(x, y) = f_l(x^k, y^k) + \langle g_{x\_ap}^{k-1}, x - x^k\rangle + \langle g_{y\_ap}^{k-1}, y - y^k\rangle.$$

Clearly,

$$l^k(x^k, y^k) = f_l(x^k, y^k). \tag{4.56}$$

Since $w_{ap}^k \in \widehat{\mathcal{W}}^k$, replacing x^k with $\hat{x}^k$ in (4.4), we have

$$\begin{aligned}
l^k(x, y) =& \langle -\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x), w_{ap}^k\rangle + b^T y + \frac{1}{2}\langle x^k, \mathcal{Q}(x^k)\rangle + \langle \mathcal{Q}(x^k), x - x^k\rangle \\
\le& \langle -\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x), w_{ap}^k\rangle + b^T y + \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle \\
\le& \max_{w\in\widehat{\mathcal{W}}^k} \langle -\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x), w\rangle + b^T y + \frac{1}{2}\langle x, \mathcal{Q}(x)\rangle \\
=& f_{\widehat{\mathcal{W}}^k}(x, y).
\end{aligned} \tag{4.57}$$

Therefore, using the expressions of $L^k(x,y,w)$ and $l^k(x,y)$, we get

$$
\begin{aligned}
0 \leq & \Delta_k \overset{(4.19)}{\leq} \eta^{k-1} \overset{(4.56)}{=} l^k(x^k,y^k) + \alpha\epsilon^k - f_{w^k}(x^k,y^k) \\
= & l^k(x^{k+1},y^{k+1}) - f_{w^k}(x^k,y^k) - \langle g_{x\_ap}^{k-1}, x^{k+1}-x^k\rangle - \langle g_{y\_ap}^{k-1}, y^{k+1}-y^k\rangle + \alpha\epsilon^k \\
\overset{(4.57)}{\leq} & f_{\widehat{\mathcal{W}}^k}(x^{k+1},y^{k+1}) - f_{w^k}(x^k,y^k) - \langle g_{x\_ap}^{k-1}, x^{k+1}-x^k\rangle - \langle g_{y\_ap}^{k-1}, y^{k+1}-y^k\rangle + \delta^{k-1} \\
\overset{(4.9)}{=} & f_{w^{k+1}}(x^{k+1},y^{k+1}) - f_{w^k}(x^k,y^k) - \langle g_{x\_ap}^{k-1}, x^{k+1}-x^k\rangle - \langle g_{y\_ap}^{k-1}, y^{k+1}-y^k\rangle + \delta^{k-1} \\
\overset{(4.5)}{\leq} & L^k(x^{k+1},y^{k+1},w^{k+1}) - L^{k-1}(x^k,y^k,w^k) - \frac{\nu_k}{2}(\|x^{k+1}-\hat{x}\|^2_{\mathcal{M}_{x,k}} + \|y^{k+1}-\hat{y}\|^2_{\mathbf{M}_{y,k}}) \\
& + \frac{\nu_{k-1}}{2}(\|x^k-\hat{x}\|^2_{\mathcal{M}_{x,k-1}} + \|y^k-\hat{y}\|^2_{\mathbf{M}_{y,k-1}}) + \|b-\mathcal{A}(w_{ap}^k)\|_2\|y^k-y^{k+1}\|_2 \\
& + \|x^{k+1}-x^k\|_F\|\mathcal{Q}(w_{ap}^k-x^k)\|_F + \delta^{k-1},
\end{aligned}
$$

Since $w^k, w_{ap}^k \in \alpha\mathcal{W}$ for any $k \geq N$, $\{\|x^k\|_F\}$, $\{\|y^k\|_2\}$, $\{\|w^k\|_F\}$ and $\{\|w_{ap}^k\|_F\}$ are bounded.

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

By (4.47) and (4.48), we get

$$x^k - \hat{x} = \frac{1}{\nu_{k-1}}\mathcal{M}_{x,k-1}^{-1}(\mathcal{Q}(w^k - x^k)) \to \mathbf{0} \tag{4.58}$$

$$y^k - \hat{y} = \frac{1}{\nu_{k-1}}\mathbf{M}_{y,k-1}^{-1}(\mathcal{A}(w^k) - b) \to \mathbf{0} \tag{4.59}$$

as $k \to \infty$ since both $\{\|w^k\|_F\}$ and $\{\|x^k\|_F\}$ are bounded, which proves the statement (4.53). Moreover, in view of (4.47) and the definition of $\mathcal{T}_{k-1}$, we have

$$\mathcal{Q}(x^k - \hat{x}) + \nu_{k-1}\mathcal{M}_{x,k-1}(x^k - \hat{x}) = \mathcal{Q}(w^k - \hat{x}),$$

whence

$$(x^k - \hat{x}) = \frac{1}{\nu_{k-1}}\mathcal{M}_{x,k-1}^{-1}\mathcal{Q}(w^k - x^k).$$

Therefore, together with (4.48), it holds

$$
\begin{aligned}
& \frac{\nu_{k-1}}{2}(\|x^k-\hat{x}\|^2_{\mathcal{M}_{x,k-1}} + \|y^k-\hat{y}\|^2_{\mathbf{M}_{y,k-1}}) - \frac{\nu_k}{2}(\|x^{k+1}-\hat{x}\|^2_{\mathcal{M}_{x,k}} + \|y^{k+1}-\hat{y}\|^2_{\mathbf{M}_{y,k}}) \\
\leq & \frac{\nu_{k-1}}{2}(\|x^k-\hat{x}\|^2_{\mathcal{M}_{x,k-1}} + \|y^k-\hat{y}\|^2_{\mathbf{M}_{y,k-1}})
\end{aligned}
$$

$$=\frac{1}{2\nu_{k-1}}(\|\mathcal{Q}(w^k-x^k)\|^2_{\mathcal{M}^{-1}_{x,k-1}}+\|\mathcal{A}(w^k)-b\|^2_{\mathbf{M}^{-1}_{y,k-1}}).$$

Using the above inequality, we further bound

$$\begin{aligned}
&\eta_{k-1}\\
\le& L^k(x^{k+1},y^{k+1},w^{k+1})-L^{k-1}(x^k,y^k,w^k)+\frac{1}{2\nu_{k-1}}(\|\mathcal{Q}(w^k-x^k)\|^2_{\mathcal{M}^{-1}_{x,k-1}}+\|\mathcal{A}(w^k)-b\|^2_{\mathbf{M}^{-1}_{y,k-1}})\\
&+\|b-\mathcal{A}(w^k_{ap})\|_2\|y^k-y^{k+1}\|_2+\|x^{k+1}-x^k\|_F\|\mathcal{Q}(w^k_{ap}-x^k)\|_F+\delta^{k-1},
\end{aligned}$$

The boundedness of $\{\|x\|_F\}$ and $\{\|w^k\|_F\}$ and $\lim_{k\to\infty}\nu_k=\infty$ result in

$$\lim_{k\to\infty}\frac{1}{2\nu_{k-1}}(\|\mathcal{Q}(w^k-x^k)\|^2_{\mathcal{M}^{-1}_{x,k-1}}+\|\mathcal{A}(w^k)-b\|^2_{\mathbf{M}^{-1}_{y,k-1}})=0.$$

Therefore, together with (4.51), (4.49) and $\delta^k\to 0$, it yields $\lim\limits_{k\to\infty}\eta^{k-1}=0$, whence $\lim\limits_{k\to\infty}\Delta_k=0$. Combining this with (4.53), we obtain (4.52).

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

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

$$\begin{aligned}
e_v^{k-1}\overset{(4.42)}{=}&f_u(\hat{x}^{k-1},\hat{y}^{k-1})-f_{w^k_{ap}}(\hat{x}^{k-1},\hat{y}^{k-1})\\
=&\alpha(\hat{\lambda}^{k-1}_{ap}+\hat{\epsilon}^{k-1})-\langle -\hat{s}-\mathcal{A}^*(\hat{y})-\mathcal{Q}(\hat{x}),w^k_{ap}\rangle\\
\ge&\alpha\lambda_1(-\hat{s}-\mathcal{A}^*(\hat{y})-\mathcal{Q}(\hat{x}))-\langle -\hat{s}-\mathcal{A}^*(\hat{y})-\mathcal{Q}(\hat{x}),w^k_{ap}\rangle\ge 0.
\end{aligned}$$

Moreover, by step 6,

$$\alpha\hat{\epsilon}^k\le\min\{\delta^{k-1},10\delta^k\},\tag{4.60}$$

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.

Clearly, $\epsilon^k\to 0$ if $\delta^k\to 0$.

By the definition of the Ritz-type triplet for the largest eigenvalue $\lambda_1(-\hat{s}-\mathcal{A}^*(y^k)-\mathcal{Q}(x^k))$, we see that

$$\alpha\lambda^k_{ap}=\langle -\hat{s}-\mathcal{A}^*(y^k)-\mathcal{Q}(x^k),w^k_{ap}\rangle.\tag{4.61}$$

Consequently, in combination with (4.58), (4.59) and $\delta^k \to 0$, we have

$$\begin{aligned}
e_v^{k-1} &\overset{(4.42),(3.14)}{\leq} \alpha\lambda_1(-\hat{s} - \mathcal{A}^*(\hat{y}) - \mathcal{Q}(\hat{x})) - \langle -\hat{s} - \mathcal{A}^*(\hat{y}) - \mathcal{Q}(\hat{x}), w_{ap}^k \rangle + \alpha\hat{\epsilon}^{k-1} \\
&= \alpha\lambda_1(-\hat{s} - \mathcal{A}^*(\hat{y}) - \mathcal{Q}(\hat{x})) - \langle -\hat{s} - \mathcal{A}^*(y^k) - \mathcal{Q}(x^k), w_{ap}^k \rangle + \alpha\hat{\epsilon}^{k-1} \\
&\quad - \langle y^k - \hat{y}, \mathcal{A}(w_{ap}^k) \rangle - \langle \mathcal{Q}(x^k - \hat{x}), w_{ap}^k \rangle \\
&\overset{(4.61)}{=} \alpha\lambda_1(-\hat{s} - \mathcal{A}^*(\hat{y}) - \mathcal{Q}(\hat{x})) - \alpha(\lambda_{ap}^k + \epsilon^k) + \alpha\epsilon^k + \alpha\hat{\epsilon}^{k-1} \\
&\quad - \langle y^k - \hat{y}, \mathcal{A}(w_{ap}^k) \rangle - \langle \mathcal{Q}(x^k - \hat{x}), w_{ap}^k \rangle \\
&\overset{\text{Proposition 3.4}}{\leq} \alpha\lambda_1(-\hat{s} - \mathcal{A}^*(\hat{y}) - \mathcal{Q}(\hat{x})) - \alpha\lambda_1(-\hat{s} - \mathcal{A}^*(y^k) - \mathcal{Q}(x^k)) + \alpha\epsilon^k + \alpha\hat{\epsilon}^{k-1} \\
&\quad - \langle y^k - \hat{y}, \mathcal{A}(w_{ap}^k) \rangle - \langle \mathcal{Q}(x^k - \hat{x}), w_{ap}^k \rangle \\
&\leq \alpha((L\|\mathcal{A}^*\|_2 + \|\mathcal{A}(w_{ap}^k)\|_2)\|y^k - \hat{y}\|_2 + (L\|\mathcal{Q}\|_2 + \|\mathcal{Q}(\mathcal{A}(w_{ap}^k))\|_F)\|x^k - \hat{x}\|_F + \epsilon^k + \hat{\epsilon}^{k-1}) \\
&\to 0
\end{aligned}$$

as $k \to \infty$, where L is the Lipschitz constant of $\lambda_1(\cdot)$.

By step 6, the above expression yields $\theta_v^k \to 0$. Therefore

$$\vartheta^{k-1} + \alpha\hat{\epsilon}^{k-1} \to 0.$$

Subsequently, by (4.38), it holds

$$\nu_{k-1} l_d^{k-1} + \tau^{k-1} \to 0 \text{ as } k \to \infty,$$

whence

$$b - \mathcal{A}(w^k) = -\nu_{k-1}\mathbf{M}_{y,k-1}(d_y^{k-1}) \to \mathbf{0}$$

and

$$\mathcal{Q}(x^k - w^k) = -\nu_{k-1}\mathcal{M}_{x,k-1}(d_x^{k-1}) \to \mathbf{0} \text{ as } k \to \infty$$

from (4.58), (4.59) and (4.13), which proves the statement (4.54).

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

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

Since $\nu^k \le \bar{\nu}$ for any $k \ge N$, with the help of (1.4), we further bound

$$
\begin{aligned}
\eta_{k-1} \le & L^k(x^{k+1}, y^{k+1}, w^{k+1}) - L^{k-1}(x^k, y^k, w^k) - \frac{\nu_k}{2}(\|x^{k+1} - \hat{x}\|^2_{\mathcal{M}_{x,k}} + \|y^{k+1} - \hat{y}\|^2_{\mathbf{M}_{y,k}}) \\
& + \frac{\nu_k}{2}(\|x^k - \hat{x}\|^2_{\mathcal{M}_{x,k}} + \|y^k - \hat{y}\|^2_{\mathbf{M}_{y,k}}) + \|b - \mathcal{A}(w^k_{ap})\|_2 \|y^k - y^{k+1}\|_2 \\
& + \|x^{k+1} - x^k\|_F \|\mathcal{Q}(w^k_{ap} - x^k)\|_F + \delta^{k-1} \\
= & L^k(x^{k+1}, y^{k+1}, w^{k+1}) - L^{k-1}(x^k, y^k, w^k) + \nu_k \langle x^k - x^{k+1}, \mathcal{M}_{x,k}(x^k - \hat{x})\rangle \\
& + \nu_k \langle y^k - y^{k+1}, \mathbf{M}_{y,k}(y^k - \hat{y})\rangle - \frac{\nu_k}{2}\|x^k - x^{k+1}\|^2_{\mathcal{M}_{x,k}} - \frac{\nu_k}{2}\|y^k - y^{k+1}\|^2_{\mathbf{M}_{y,k}} \\
& + \|b - \mathcal{A}(w^k_{ap})\|_2 \|y^k - y^{k+1}\|_2 + \|x^{k+1} - x^k\|_F \|\mathcal{Q}(w^k_{ap} - x^k)\|_F + \delta^{k-1} \\
\le & L^k(x^{k+1}, y^{k+1}, w^{k+1}) - L^{k-1}(x^k, y^k, w^k) \\
& + (\nu_k \|\mathbf{M}_{y,k}(y^k - \hat{y})\|_2 + \|b - \mathcal{A}(w^k_{ap})\|_2)\|y^k - y^{k+1}\|_2 \\
& + (\nu_k \|\mathcal{M}_{x,k}(x^k - \hat{x})\|_F + \|\mathcal{Q}(w^k_{ap} - x^k)\|_F)\|x^{k+1} - x^k\|_F + \delta^{k-1} \\
\le & L^k(x^{k+1}, y^{k+1}, w^{k+1}) - L^{k-1}(x^k, y^k, w^k) \\
& + (\bar{\nu} \|\mathbf{M}_{y,k}(y^k - \hat{y})\|_2 + \|b - \mathcal{A}(w^k_{ap})\|_2)\|y^k - y^{k+1}\|_2 \\
& + (\bar{\nu} \|\mathcal{M}_{x,k}(x^k - \hat{x})\|_F + \|\mathcal{Q}(w^k_{ap} - x^k)\|_F)\|x^{k+1} - x^k\|_F + \delta^{k-1}.
\end{aligned}
$$

Therefore, together with (4.51), (4.49), $\lim_{k\to\infty} \delta^k = 0$ and the boundedness of $\{\|x^k\|_F\}$, $\{\|y^k\|_2\}$ and $\{\|w^k_{ap}\|_F\}$, it follows $\lim\limits_{k\to\infty} \eta^{k-1} = 0$, whence $\lim\limits_{k\to\infty} \Delta_k = 0$.

If no descent steps occur, then

$$
f_l(\hat{x}^{k-1}, \hat{y}^{k-1}) - f_l(x^k, y^k) < m_L(f_l(\hat{x}^{k-1}, \hat{y}^{k-1}) - f_{w^k}(x^k, y^k)),
$$

thus for $k > N$, it follows

$$
\begin{aligned}
\Delta_k + \delta^{k-1} \ge \Delta_k + \alpha\varepsilon^k \ge & f_u(x^k, y^k) - f_{w^k}(x^k, y^k) \ge f_l(x^k, y^k) - f_{w^k}(x^k, y^k) \\
> & (1 - m_L)(f_l(\hat{x}^{k-1}, \hat{y}^{k-1}) - f_{w^k}(x^k, y^k)) \\
\ge & (1 - m_L)(f(\hat{x}, \hat{y}) - \alpha\hat{\epsilon}^{k-1} - f_{w^k}(x^k, y^k)) \\
\overset{(4.27)}{\ge} & -\alpha\hat{\epsilon}^{k-1}(1 - m_L) \ge -\delta^{k-2}(1 - m_L).
\end{aligned}
$$

whence (4.52) holds. Thus the proof is complete by Lemma 4.2. □

Lemma 4.4. *If a sequence $\{(x^k, y^k)\}_{k=N}^{\infty}$ possesses the properties (4.52) , (4.53),* *(4.54) and (4.14), then* $(\hat{x}, \hat{y}) \in arg\min_{x,y} f(x, y)$.

Proof. By means of (4.14), $(\mathcal{Q}(x^k - w^k), b - \mathcal{A}(w^k)) \in \partial f_{\widehat{\mathcal{W}}^{k-1}}(x^k, y^k)$, we obtain that, for any $(x, y) \in \mathbb{V} \times \mathbb{R}^m$,

$$\begin{aligned} &f(x, y) \geq f_{\widehat{\mathcal{W}}^{k-1}}(x, y) \\ &\geq f_{w^k}(x^k, y^k) + \langle \mathcal{Q}(x^k - w^k), x - x^k \rangle + \langle b - \mathcal{A}(w^k), y - y^k \rangle \\ &\xrightarrow[(4.53),(4.54)]{(4.52)} f(\hat{x}, \hat{y}) \end{aligned}$$

as $k \to \infty$, whence $(\hat{x}, \hat{y}) \in arg\min_{x,y} f(x, y)$. □

Proposition 4.6. *Let* $\{(x^k, y^k)\}_{k=N}^{\infty}$ *be specified by (4.47) and (4.48). If Assumption* **A3** *is satisfied and* $\delta^k \to 0$, *then* $(x^k, y^k) \to (\hat{x}, \hat{y}) \in arg\min_{x,y} f(x, y)$ *and* $\lim_{k\to\infty} \varpi^k = 0$.

Proof. Putting together Lemma 4.1, Lemma 4.2, Lemma 4.3 and Lemma 4.4, we get $(x^k, y^k) \to (\hat{x}, \hat{y}) \in arg\min_{x,y} f(x, y)$. By virtue of (4.13), $(x^k, y^k) \to (\hat{x}, \hat{y})$ results in $(g_x^{k-1}, g_y^{k-1}) \to (\mathbf{0}, \mathbf{0})$. Hence, in combination with the boundedness of ϱ_k (see (4.22) for its definition), we obtain

$$\varrho_{k-1}\nu_{k-1}\sqrt{l_d^{k-1}} = \varrho_{k-1}(\|\mathcal{M}_{x,k-1}^{-1/2}(g_x^{k-1})\|_F^2 + \|\mathbf{M}_{y,k-1}^{-1/2} g_y^{k-1}\|_2^2)^{1/2} \to 0 \text{ as } k \to \infty.$$

Moreover,

$$0 \leq \min\{\eta^{k-1}, \tau^{k-1}\} \leq \eta^{k-1} \to 0$$

as $k \to \infty$ by the proof of Lemma 4.3, hence $\lim_{k\to\infty} \varpi^{k-1} = 0$. □

We now focus on the case that the number of descent steps approaches infinity. To guarantee the convergence of Algorithm 4.4 in this setting, we take the following assumption.

(A4) $\alpha\lambda_{ap}^{k+1} \geq \langle w^{k+1}, -\hat{s} - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(x^{k+1}) \rangle$ for all k.

Remark 4.5. *Let* $-\hat{s} - \mathcal{A}^*(y^{k+1}) - \mathcal{Q}(x^{k+1})$ *denote by* z^{k+1}. *We will see that the only property needed from the Ritz-type triplet for* $\lambda_1(z^{k+1})$ *is* $\lambda_{ap}^{k+1} \leq \lambda_1(z^{k+1}) \leq \lambda_{ap}^{k+1} + \epsilon^{k+1}$, *which is established by Proposition 3.4. At iteration* k, *after step 3, if* $\alpha\lambda_{ap}^{k+1} <$

$\langle w^{k+1}, z^{k+1}\rangle$, *then* $\langle w^{k+1}, z^{k+1}\rangle \leq \alpha\lambda_1(z^{k+1}) \leq \alpha(\lambda_{ap}^{k+1} + \epsilon^{k+1}) < \langle w^{k+1}, z^{k+1}\rangle + \alpha\epsilon^{k+1}$. *In addition, it follows from Proposition 4.5 that* $w^{k+1} \in \widehat{\mathcal{W}}^{k+1}$. *Thus,* $f_l(x^{k+1}, y^{k+1}) \leq f(x^{k+1}, y^{k+1}) \leq f_l(x^{k+1}, y^{k+1}) + \alpha\epsilon^{k+1}$ *and* $w_{ap}^{k+1} \in \widehat{\mathcal{W}}^{k+1}$, *which are needed to guarantee convergence, are satisfied by setting*

$$\lambda_{ap}^{k+1} := \frac{1}{\alpha}\langle w^{k+1}, z^{k+1}\rangle \text{ and } w_{ap}^{k+1} := w^{k+1}. \tag{4.62}$$

In view of (4.62), *Assumption* **A4** *holds.*

The preceding result is with respect to convergence of Algorithm 4.4 when the number s of descent steps is bounded. It remains to analyze the case of an unbounded s. Over the course of executing the algorithm, it may happen that null steps appear between two consecutive descent steps. Since we are investigating the case that the number of descent steps approaches infinity, for the sake of convenience, we discard all null steps. In other words, we focus on the situation $(\hat{x}^k, \hat{y}^k) = (x^k, y^k)$. Thus we relabel the remaining iterates and the corresponding w^k with a new index h, and assume that, for any h,

$$x^{h+1} = x_{min}^h(w^{h+1}) = x^h + \mathcal{T}_h^{-1}(\mathcal{Q}(w^{h+1} - x^h)), \tag{4.63}$$

$$y^{h+1} = y_{min}^h(w^{h+1}) = y^h + \frac{1}{\nu_h}\mathbf{M}_{y,h}^{-1}(\mathcal{A}(w^{h+1}) - b), \tag{4.64}$$

$$f_l(x^h, y^h) - f_l(x^{h+1}, y^{h+1}) \geq m_L(f_l(x^h, y^h) - f_{w^{h+1}}(x^{h+1}, y^{h+1})). \tag{4.65}$$

By Assumption **A4**, it holds

$$f_l(x^h, y^h) - m_L(f_l(x^h, y^h) - f_{w^{h+1}}(x^{h+1}, y^{h+1})) \geq f_l(x^{h+1}, y^{h+1}) \geq f_{w^{h+1}}(x^{h+1}, y^{h+1}),$$

whence

$$(1 - m_L)(f_l(x^h, y^h) - f_{w^{h+1}}(x^{h+1}, y^{h+1})) \geq 0.$$

Therefore,

$$f_l(x^h, y^h) - f_l(x^{h+1}, y^{h+1}) \geq 0. \tag{4.66}$$

Owing to (4.66), the sequence $\{f_l(x^h, y^h)\}_{h=1}^{\infty}$ is nonincreasing. At this moment, we make the following assumption:

$$\text{There is some } (\tilde{x},\tilde{y}) \in \mathbb{V}\times\mathbb{R}^m \text{ such that } f_l(x^h,y^h) \geq f(\tilde{x},\tilde{y}) \text{ for all } h. \tag{4.67}$$

Proposition 4.7. *Let $\{(x^h,y^h)\}_{h=1}^{\infty}$ be specified by (4.63) and (4.64). Assume that (4.67), $\delta^h \downarrow 0$, and Assumptions* **A3** *and* **A4** *are satisfied. Then (x^h,y^h) converges to a minimizer of $f(x,y)$, and $\lim_{h\to\infty}\varpi^h = 0$.*

Proof. We first prove that $\{(x^h,y^h)\}_{h=1}^{\infty}$ is bounded.

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

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

By step 5, for the descent steps, it follows $\nu_{h-1}\geq\nu_h\geq\nu_{min}$ for any h, in combination with (1.3), we have

$$\begin{aligned}
d_w^{h+1} =&\nu_h(\|\tilde{x}-x^h+x^h-x^{h+1}\|^2_{\mathcal{M}_{x,h}} + \|\tilde{y}-y^h+y^h-y^{h+1}\|^2_{\mathbf{M}_{y,h}})\\
\leq&\nu_h(\|\tilde{x}-x^h\|^2_{\mathcal{M}_{x,h}} + 2\langle\tilde{x}-x^h,\mathcal{M}_{x,h}(x^h-x^{h+1})\rangle + 2\langle x^h-x^{h+1},\mathcal{M}_{x,h}(x^h-x^{h+1})\rangle\\
&+\|\tilde{y}-y^h\|^2_{\mathbf{M}_{y,h}} + 2\langle\tilde{y}-y^h,\mathbf{M}_{y,h}(y^h-y^{h+1})\rangle + 2\langle y^h-y^{h+1},\mathbf{M}_{y,h}(y^h-y^{h+1})\rangle)\\
=&\nu_h(\|\tilde{x}-x^h\|^2_{\mathcal{M}_{x,h}} + \|\tilde{y}-y^h\|^2_{\mathbf{M}_{y,h}}) + 2(\langle\mathcal{Q}(x^{h+1}-w^{h+1}),\tilde{x}-x^{h+1}\rangle\\
&+\langle\tilde{y}-y^{h+1},b-\mathcal{A}(w^{h+1})\rangle)\\
\leq& d_w^h + 2(\langle\mathcal{Q}(x^{h+1}-w^{h+1}),\tilde{x}-x^{h+1}\rangle + \langle\tilde{y}-y^{h+1},b-\mathcal{A}(w^{h+1})\rangle).
\end{aligned}$$

Recall that

$$(g_x^h,g_y^h) = (\mathcal{Q}(x^{h+1}-w^{h+1}),b-\mathcal{A}(w^{h+1})) \in \partial f_{\widehat{\mathcal{W}}^h}(x^{h+1},y^{h+1}) = \nabla f_{w^{h+1}}(x^{h+1},y^{h+1}). \tag{4.68}$$

Therefore, we obtain

$$\begin{aligned}
&\langle g_x^h,\tilde{x}-x^{h+1}\rangle + \langle g_y^h,\tilde{y}-y^{h+1}\rangle \leq f_{\widehat{\mathcal{W}}^h}(\tilde{x},\tilde{y}) - f_{w^{h+1}}(x^{h+1},y^{h+1})\\
\leq& f(\tilde{x},\tilde{y}) - f_{w^{h+1}}(x^{h+1},y^{h+1}) \overset{(4.67)}{\leq} f_l(x^h,y^h) - f_{w^{h+1}}(x^{h+1},y^{h+1}).
\end{aligned}$$

Consequently, there holds $d_w^{h+1} \leq d_w^h + \frac{2}{m_L}(f_l(x^h,y^h) - f_{w^{h+1}}(x^{h+1},y^{h+1}))$. Hence, for any fixed $H \geq 1$ and any $h > H$, we get

$$d_w^{h+1} \leq d_w^H + \frac{2}{m_L}\sum_{i=H}^{h-1}(f_l(x^i,y^i) - f_l(x^{i+1},y^{i+1})) \leq d_w^H + \frac{2}{m_L}\sum_{i=1}^{\infty}(f_l(x^i,y^i) - f_l(x^{i+1},y^{i+1})).$$

Since $\sum_{i=1}^{\infty}(f_l(x^i,y^i) - f_l(x^{i+1},y^{i+1})) = \lim_{n\to\infty}\sum_{i=1}^{n}(f_l(x^i,y^i) - f_l(x^{i+1},y^{i+1})) = \lim_{n\to\infty}(f_l(x^1,y^1) - f_l(x^{n+1},y^{n+1})) \overset{(4.67)}{\leq} f_l(x^1,y^1) - f(\tilde{x},\tilde{y})$, there holds $\sum_{i=1}^{\infty}(f_l(x^i,y^i) - f_l(x^{i+1},y^{i+1})) < \infty$, whence $\{(x^h,y^h)\}_{h=1}^{\infty}$ is bounded, and then $\{(x^h,y^h)\}_{h=1}^{\infty}$ has an accumulation point, say $(x^\star,y^\star)$, namely, there exists $\{(x^{h_j},y^{h_j})\}_{j=1}^{\infty}$ such that

$$(x^{h_j},y^{h_j}) \to (x^\star,y^\star) \text{ and } f(x^{h_j},y^{h_j}) \to f(x^\star,y^\star) \quad \text{as } j\to\infty.$$

Now we show that $(x^h,y^h) \to (x^\star,y^\star)$, whence $(g_x^h, g_y^h) \to (\mathbf{0},\mathbf{0})$ as $h\to\infty$.

If $\lim_{h\to\infty}\delta^h = 0$, then $\epsilon^h \to 0$. In combination with $f_l(x^h,y^h) \leq f(x^h,y^h) \leq f_l(x^h,y^h) + \alpha\epsilon^h$, we have $\lim_{h\to\infty} f_l(x^h,y^h) = \lim_{h\to\infty} f(x^h,y^h) = f(x^\star,y^\star)$, whence $f_l(x^h,y^h) \geq f(x^\star,y^\star)$ for all h. By the same argument as the above, we have, for any $h > H$,

$$\begin{aligned}
&\nu_h(\|x^\star - x^{h+1}\|^2_{\mathcal{M}_{x,h}} + \|y^\star - y^{h+1}\|^2_{\mathbf{M}_{y,h}}) \\
\leq&\nu_{H-1}(\|x^\star - x^H\|^2_{\mathcal{M}_{x,H-1}} + \|y^\star - y^H\|^2_{\mathbf{M}_{y,H-1}}) + 2\sum_{i=H}^{\infty}(f_l(x^i,y^i) - f_l(x^{i+1},y^{i+1})).
\end{aligned}$$

Therefore, for any arbitrary $\epsilon > 0$, one can choose large enough $\widehat{H}$, such that for any $h > \widehat{H}$, $\nu_h(\|x^\star - x^{h+1}\|^2_{\mathcal{M}_{x,h}} + \|y^\star - y^{h+1}\|^2_{\mathbf{M}_{y,h}}) < \epsilon$, whence $(x^h,y^h) \to (x^\star,y^\star)$ as $h\to\infty$. Consequently, $g_x^h = \mathcal{Q}(x^{h+1} - w^{h+1}) = \nu_h\mathcal{M}_{x,h}(x^h - x^{h+1}) \to \mathbf{0}$ and $g_y^h = b - \mathcal{A}(w^{h+1}) = \nu_h\mathbf{M}_{y,h}(y^h - y^{h+1}) \to \mathbf{0}$ as $h\to\infty$, and then

$$\lim_{h\to\infty}\varrho_h\nu_h\sqrt{l_d^h} = 0. \tag{4.69}$$

Using (4.66) and (4.67), we see that

$$\lim_{h\to\infty} f_l(x^h,y^h) = \inf_h f_l(x^h,y^h) \in (-\infty,\infty). \tag{4.70}$$

It then follows from (4.65) that

$$\lim_{h\to\infty} f_l(x^h, y^h) = \lim_{h\to\infty} f_{w^{h+1}}(x^{h+1}, y^{h+1}). \tag{4.71}$$

Note that

$$\begin{aligned} f(x,y) \geq f_{w^{h+1}}(x,y) &\overset{(4.68)}{\geq} f_{w^{h+1}}(x^{h+1}, y^{h+1}) + \langle g_x^h, x - x^{h+1}\rangle + \langle g_y^h, y - y^{h+1}\rangle \\ \geq & f_{w^{h+1}}(x^{h+1}, y^{h+1}) - \|g_x^h\|_F \|x - x^{h+1}\|_F - \|g_y^h\|_2 \|y - y^{h+1}\|_2, \ \forall\ (x,y) \in \mathbb{V} \times \mathbb{R}^m. \end{aligned}$$

Together with the boundedness of $\{(x^h, y^h)\}$, it then follows that

$$f(x,y) \geq \lim_{h\to\infty} f_{w^{h+1}}(x^{h+1}, y^{h+1}) \overset{(4.71)}{=} \lim_{h\to\infty} f_l(x^h, y^h) = f(x^\star, y^\star), \ \forall\ (x,y) \in \mathbb{V} \times \mathbb{R}^m.$$

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

$$f_l(x^h, y^h) + \alpha\epsilon^{h+1} - f_{w^{h+1}}(x^{h+1}, y^{h+1}) \overset{(4.66)}{\geq} f_u(x^{h+1}, y^{h+1}) - f_{w^{h+1}}(x^{h+1}, y^{h+1}) = \eta^{h+1} \geq 0,$$

this, together with $\lim_{h\to\infty} \epsilon^h = 0$ and (4.71), implies $\lim_{h\to\infty} \eta^h = 0$, hence $\lim_{h\to\infty} \varpi^h = 0$ by (4.69). □

.

We summarize the above discussion in the following theorem.

Theorem 4.2. *Let* $\left\{(\hat{x}^k, \hat{y}^k)\right\}_{k=1}^\infty$ *be the sequence of points generated by Algorithm 4.4. Suppose that Assumptions* **A3** *and* **A4** *are satisfied and* $\lim_{k\to\infty} \delta^k = 0$*, then either* $(\hat{x}^k, \hat{y}^k) \to (\bar{x}, \bar{y}) \in arg\min_{x,y} f(x,y)$ *or there is no minimizer and* $\lim_{k\to\infty} \|(\hat{x}^k, \hat{y}^k)\| = \infty$.

Proof. Case 1: a finite number of descent steps.

In this case, there exists iteration N, such that starting with iteration N, the descent test fails in all subsequent iterations in Algorithm 4.4. By Algorithm 4.4, $(\hat{x}^k, \hat{y}^k) = (\hat{x}^N, \hat{y}^N)$ for any $k > N$. Proposition 4.6 shows that

$$(\hat{x}^k, \hat{y}^k) \to (\hat{x}^N, \hat{y}^N) =: (\bar{x}, \bar{y}) \in arg\min_{x,y} f(x,y).$$

Clearly, $\lim_{k\to\infty} f(\hat{x}^k, \hat{y}^k) \to \inf f(x,y)$.

Case 2: the number of descent steps approaches infinity. In this case, without loss of generality, we can assume that for some K, $(\hat{x}^k, \hat{y}^k) = (x^k, y^k)$ for any $k > K$.

Subcase 1: the assumption (4.67) is satisfied. By virtue of Proposition 4.7, the desired results hold.

Subcase 2: the assumption (4.67) is invalid. Then for any $(x, y) \in \mathbb{V} \times \mathbb{R}^m$, there exists some integer $H(x, y)$ such that $f_l(x^{H(x,y)}, y^{H(x,y)}) < f(x, y)$. Thus we have $\inf f_l(x^k, y^k) \leq \inf f(x, y)$.

Since $\lim_{k\to\infty} \delta^k = 0$ implies $\lim_{k\to\infty} \epsilon^k = 0$, we have

$$\lim_{k\to\infty} f_l(x^k, y^k) = \lim_{k\to\infty} f(x^k, y^k) \geq \inf f(x, y),$$

whence $\inf f_l(x^k, y^k) = \inf f(x, y)$ as $\{f_l(x^k, y^k)\}_{k=1}^{\infty}$ is decreasing. As f is continuous, either boundedness of the sequence $\{(\hat{x}^k, \hat{y}^k)\}_{k=1}^{\infty}$ or attainment of $\inf f(x, y)$ would imply existence of a point $(\bar{x}, \bar{y})$ with $f_l(x^k, y^k) \geq f(\bar{x}, \bar{y})$ for all k, which contradict that (4.67) is invalid. Therefore, $arg \min_{x,y} f(x, y) = \emptyset$ and $\|(\hat{x}^k, \hat{y}^k)\| \to \infty$ as $k \to \infty$. □

So far, we have investigated the algorithm and its convergence for (EigForm). It is natural to ask whether the same algorithm can be applied to handle (P). Thanks to Proposition 3.1, we can relate the accumulation points of $\{w^k\}$ to the optimal solutions of (P).

Theorem 4.3. *Let $\mathcal{O}_P = arg \min\{\frac{1}{2}\langle x, \mathcal{Q}(x)\rangle + \langle \hat{s}, x\rangle \mid \mathcal{A}(x) = b, x \succeq \mathbf{0}\}$. Assume that $arg \min\limits_{x,y} f(x, y) \neq \emptyset$, Assumptions* **A1**, **A3** *and* **A4** *are satisfied and $\lim_{k\to\infty} \delta^k = 0$. Let $\{w^k\}$ be the sequence of points generated by Algorithm 4.4. Then all accumulation points of $\{w^k\}$ lie in $\mathcal{O}_P$.*

Proof. Set

$$p := \min \left\{ h(x) \mid \mathcal{A}(x) = b, x \succeq \mathbf{0} \right\},$$
$$-d := \min \left\{ g(x, y) \mid \mathcal{Q}(x) + \mathcal{A}^*(y) + \hat{s} \succeq \mathbf{0} \right\}.$$

Then Assumption **A1**, $arg \min_{x,y} f(x, y) \neq \emptyset$ and Proposition 3.1 imply $p = d$.

Let $\{(x^k, y^k)\}_{k=1}^{\infty}$ be a sequence of points generated by Algorithm 4.4, which contains a subsequence $\{(\hat{x}^k, \hat{y}^k)\}_{k=1}^{\infty}$.

Case 1: If k is finite with $k = K$ on termination, then by Proposition 4.3, it holds $(x^{K+1}, y^{K+1}) = (\hat{x}^K, \hat{y}^K)$ and (x^{K+1}, y^{K+1}) is optimal, thus

$$\mathcal{Q}(x^{K+1} - w^{K+1}) = \mathbf{0} \text{ and } b - \mathcal{A}(w^{K+1}) = \mathbf{0}. \tag{4.72}$$

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

On the other hand, we have

$$\begin{aligned}
&-h(w^{K+1}) = -\frac{1}{2}\langle w^{K+1}, \mathcal{Q}(w^{K+1})\rangle - \langle \hat{s}, w^{K+1}\rangle \overset{(4.72)}{=} -\frac{1}{2}\langle w^{K+1}, \mathcal{Q}(x^{K+1})\rangle - \langle \hat{s}, w^{K+1}\rangle \\
&\overset{(4.72)}{=} \langle -\hat{s} - \mathcal{A}^*(y^{K+1}) - \mathcal{Q}(x^{K+1}), w^{K+1}\rangle + b^T y^{K+1} + \frac{1}{2}\langle x^{K+1}, \mathcal{Q}(x^{K+1})\rangle \\
&\overset{\textit{Proposition 4.3}}{=} f(\hat{x}^K, \hat{y}^K) \overset{\textit{Proposition 3.1}}{=} g(\hat{x}^K, \hat{y}^K) = -d,
\end{aligned}$$

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

Case 2: The number of null steps approaches infinity.

In this case, there exists iteration N, such that starting with iteration N, the descent test fails in all subsequent iterations in Algorithm 4.4. Let $\{(x^k, y^k)\}_{k=N}^{\infty}$ be specified by (4.47) and (4.48). Since $w^{k+1} \in \widehat{\mathcal{W}}^k \subseteq \alpha\mathcal{W}$ and $\alpha\mathcal{W}$ is compact, $\{w^k\}$ has accumulation points and they are contained in $\alpha\mathcal{W}$.

Assume that a subsequence $\{w^{k_j}\}$ converges a point, say $w_\star \in \alpha\mathcal{W}$. By Lemma 4.3, we see that

$$x^{k_j} \to \hat{x}^N, y^{k_j} \to \hat{y}^N$$

and

$$f_{w^{k_j}}(x^{k_j}, y^{k_j}) \to f(\hat{x}^N, \hat{y}^N),\ b - \mathcal{A}(w^{k_j}) \to \mathbf{0},\ \mathcal{Q}(\hat{x}^N - w^{k_j}) \to \mathbf{0},$$

whence

$$\mathcal{Q}(\hat{x}^N - w_\star) = \mathbf{0},\ b - \mathcal{A}(w_\star) = \mathbf{0}. \tag{4.73}$$

Therefore, we have

$$\begin{aligned}
&f_{w^{k_j}}(x^{k_j}, y^{k_j}) = \langle -\hat{s}, w^{k_j}\rangle + \langle b - \mathcal{A}(w^{k_j}), y^{k_j}\rangle + \langle x^{k_j}, \mathcal{Q}(\hat{x}^N - w^{k_j})\rangle - \frac{1}{2}\langle \hat{x}^N, \mathcal{Q}(\hat{x}^N)\rangle \\
&+ \frac{1}{2}\langle x^{k_j} - \hat{x}^N, \mathcal{Q}(x^{k_j} - \hat{x}^N)\rangle \to \langle -\hat{s}, w_\star\rangle - \frac{1}{2}\langle \hat{x}^N, \mathcal{Q}(\hat{x}^N)\rangle
\end{aligned}$$

$$= \langle -\hat{s}, w_\star \rangle - \frac{1}{2}\langle w_\star, \mathcal{Q}(w_\star)\rangle = -h(w_\star) = f(\hat{x}^N, \hat{y}^N),$$

and hence

$$-h(w_\star) = f(\hat{x}^N, \hat{y}^N) \geq -d = -p \geq -h(x)$$

for any feasible solution x for (P), where the first inequality results from Proposition 3.1. Together with (4.73), the desired result holds.

Case 3: The number of descent steps approaches infinity. In this case, Let $\{(x^h, y^h)\}_{h=1}^{\infty}$ be specified by (4.63) and (4.64).

Since $arg\min_{x,y} f(x, y) \neq \emptyset$, by the proof of Theorem 4.2, we see that (4.67) is satisfied. By Proposition 4.7, following the same approach in Case 2, we get the desired result. □

Chapter 5

Approximate solutions and a Lipschitzian error bound for the eigenvalue minimization problem

In most cases, it may be tremendously expensive to pursue an exact optimal solution. It is usually hard to find such solutions to many mathematical programs. Meanwhile, most numerical methods attempting to solve global optimization problems only yield approximately optimal solutions. Therefore, it is sensible to search for approximately optimal solutions.

In this chapter, we explore the limiting behavior of a sequence of trial points in the context that the error in eigenvalue approximation approaches a positive scalar δ. It turns out the proposed method either asymptotically finds points that are δ-optimal solutions (see Section 5.1 for its definition), or estimates asymptotically the optimal value of (Eigform) with an accuracy δ. We also provide a strategy to manage the eigenvalue tolerance δ^k to ensure convergence. In the setting of inexact computations, we would like to know how good these solutions are. In optimization theory, this is often reduced to measure the distance to the solution set from a certain given point. To this end, we will also study a Lipschitzian error bound for the eigenvalue minimization problem (Eigform).

5.1 Approximate solutions of the eigenvalue minimization problem

In this section, we restrict our discussion to the setting where the error in eigenvalue approximation approaches a positive constant. We will show that the proposed method asymptotically finds points that are approximate solutions for the reformulated problem (Eigform) when the number of null steps approaches infinity; it asymptotically estimates the optimal value of (Eigform) if otherwise. Meanwhile, our proposed method produces a sequence of points as a byproduct, whose accumulation point is an approximate solution for the CQSCP (P).

At this point, we need to recall the definitions of approximate solutions in [40].

Definition 5.1. *Let* $\epsilon, \epsilon_s > 0$*. The point* $(x^\star, y^\star) \in \mathbb{V} \times \mathbb{R}^m$ *is called an* ϵ**-**ϵ_s**-optimal** *solution of (Eigform) if*

$$f(x^\star, y^\star) \leq f(x, y) + \epsilon(\|x - x^\star\|_F^2 + \|y - y^\star\|_2^2)^{1/2} + (\epsilon + \epsilon_s), \text{ for all } (x, y) \in \mathbb{V} \times \mathbb{R}^m.$$

Hence, $\epsilon = 0$ implies ϵ_s**-optimality** of $(x^\star, y^\star)$ in the sense of [63], while $\epsilon_s = 0$ yields ϵ**-optimality** of $(x^\star, y^\star)$ in the sense of [91]. Henceforth we call them ϵ_s^{F1}**-optimal** solution and ϵ^{F2}**-optimal** solution, respectively. Plainly, by virtue of (4.23) and (4.24), if $\varpi^k \leq \epsilon$, then either $(\hat{x}^k, \hat{y}^k)$ 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 4.4, 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, practical implementations may only allow for $\delta^k \downarrow \delta > 0$. Moreover, Algorithm 4.4 expect to preserve the appealing convergence properties in this setting. To this end, we need one more assumption.

(A5) $\alpha\hat{\lambda}_{ap}^k \geq \langle w^{k+1}, -\hat{s} - \mathcal{A}^*(\hat{y}^k) - \mathcal{Q}(\hat{x}^k)\rangle$ for all k.

Remark 5.1. *Let* $\hat{z}^k$ *denote* $-\hat{s} - \mathcal{A}^*(\hat{y}^k) - \mathcal{Q}(\hat{x}^k)$*. After step 2, if* $\alpha\hat{\lambda}_{ap}^k < \langle w^{k+1}, \hat{z}^k\rangle$*, using the same scheme provided by Remark 4.5, we can ensure that Assumption* **A5** *holds by setting* $\hat{\lambda}_{ap}^k := \frac{1}{\alpha}\langle w^{k+1}, \hat{z}^k\rangle$*, which results in the desired property that* $f_l(\hat{x}^k, \hat{y}^k) \leq f(\hat{x}^k, \hat{y}^k) \leq f_l(\hat{x}^k, \hat{y}^k) + \alpha\hat{\epsilon}^k$*.*

The next proposition is the counterpart of Proposition 4.6, and its proof relies on the following lemma.

Lemma 5.1. *Let* $\{(x^k, y^k)\}_{k=N}^{\infty}$ *be specified by (4.47) and (4.48). If* $\delta^k \downarrow \delta > 0$*, Assumptions* **A3** *and* **A4** *are satisfied, and we choose* $\nu_{k+1} \in [\nu_k, \nu_{max}]$ *at step 6 of Algorithm 4.4, where* $\nu_{max} > 0$ *is fixed. then*

$$\lim_{k\to\infty} (f_l(x^k, y^k) - f_{w^k}(x^k, y^k)) = 0.$$

Proof. The proof is almost identical to the one of (4.55) in Lemma 4.3. The only difference is the expression Δ_k which replaced by $f_l(x^k, y^k) - f_{w^k}(x^k, y^k)$, whose quantity is nonnegative by Assumption **A4**. □

Proposition 5.1. *Let* $\{(x^k, y^k)\}_{k=N}^{\infty}$ *be specified by (4.47) and (4.48). If* $\delta^k \downarrow \delta > 0$*, Assumptions* **A3**, **A4** *and* **A5** *are satisfied, and we choose* $\nu_{k+1} \in [\nu_k, \nu_{max}]$ *at step 6 of Algorithm 4.4, where* $\nu_{max} > 0$ *is fixed. Then* $(\hat{x}^k, \hat{y}^k) \to (\hat{x}, \hat{y})$*, which is a* δ^{F_1}*-optimal solution of (*Eigform*).*

Proof. Obviously, $(\hat{x}^k, \hat{y}^k) \to (\hat{x}, \hat{y})$ as $k \to \infty$.

If no descent steps occur starting with iteration N, for all $k > N$, we have $f_l(\hat{x}^{k-1}, \hat{y}^{k-1}) - f_l(x^k, y^k) < m_L(f_l(\hat{x}^{k-1}, \hat{y}^{k-1}) - f_{w^k}(x^k, y^k))$, namely,

$$(1 - m_L)(f_l(\hat{x}^{k-1}, \hat{y}^{k-1}) - f_{w^k}(x^k, y^k)) < f_l(x^k, y^k) - f_{w^k}(x^k, y^k), \tag{5.1}$$

whence

$$\overline{\lim}_{k\to\infty}(f_l(x^k, y^k) - f_{w^k}(x^k, y^k)) \leq 0 \tag{5.2}$$

by Lemma 5.1.

From Assumption **A5**, we have

$$f_l(\hat{x}^{k-1}, \hat{y}^{k-1}) \geq f_{w^k}(\hat{x}^{k-1}, \hat{y}^{k-1}) \text{ for all } k > N. \tag{5.3}$$

In combination with (4.26), (4.47), (4.48) and $(\hat{x}^{k-1}, \hat{y}^{k-1}) = (\hat{x}, \hat{y})$, we get, for all

$k > N,\ f_{w^k}(\hat{x},\hat{y}) = L^{k-1}(\hat{x},\hat{y},w^k) \geq L^{k-1}(x^k,y^k,w^k)$, whence

$$\begin{aligned}\nu_{min}(\|x^k-\hat{x}\|^2_{\mathcal{M}_{x,k-1}} + \|y^k-\hat{y}\|^2_{\mathbf{M}_{y,k-1}}) &\leq f_{w^k}(\hat{x},\hat{y}) - f_{w^k}(x^k,y^k)\\ &\overset{(5.3)}{\leq} f_l(\hat{x}^{k-1},\hat{y}^{k-1}) - f_{w^k}(x^k,y^k) \overset{(5.1)}{\leq} \frac{1}{1-m_L}(f_l(x^k,y^k)-f_{w^k}(x^k,y^k)).\end{aligned}$$

It then follows from Assumption **A3** and Lemma 5.1 that

$$\lim_{k\to\infty}\|x^k-\hat{x}\|_F = 0 \text{ and } \lim_{k\to\infty}\|y^k-\hat{y}\|_2 = 0,$$

whence

$$\lim_{k\to\infty}(g_x^{k-1},g_y^{k-1}) = (\mathbf{0},\mathbf{0}). \tag{5.4}$$

Consequently, recalling $(g_x^{k-1},g_y^{k-1}) = \nabla f_{w^k}(x^k,y^k)$, we have for $k > N$,

$$\begin{aligned}f_l(\hat{x}^{k-1},\hat{y}^{k-1}) - f_{w^k}(x^k,y^k) &\geq f_l(\hat{x}^{k-1},\hat{y}^{k-1}) - f(x,y) + \langle g_x^{k-1}, x-x^k\rangle + \langle g_y^{k-1}, y-y^k\rangle\\ &\overset{(4.60)}{\geq} f(\hat{x},\hat{y}) - \delta^{k-2} - f(x,y) + \langle g_x^{k-1}, x-x^k\rangle + \langle g_y^{k-1}, y-y^k\rangle.\end{aligned}$$

In combination with (5.4), (5.2) and $\delta^k \downarrow \delta$, we have $f(x,y) \geq f(\hat{x},\hat{y}) - \delta$ for any $(x,y) \in \mathbb{V}\times\mathbb{R}^m$. □

The following proposition is the counterpart of Proposition 4.7.

Proposition 5.2. *Let $\{(x^h,y^h)\}_{h=1}^\infty$ be specified by (4.63) and (4.64). Assume that (4.67), $\delta^k \downarrow \delta > 0$ and Assumptions **A3**, **A4** and **A5** are satisfied. Then each accumulation point of $\{(x^h,y^h)\}_{h=1}^\infty$ is a δ^{F_1}-optimal solution of (Eigform). If, in addition, $\sum_{h=1}^\infty \epsilon^h < \infty$, then $(x^h,y^h) \to (x^\star,y^\star) \in arg\min_{x,y} f(x,y)$.*

Proof. The proof is almost identical to that of Proposition 4.7. The only difference is the proof of $\lim_{h\to\infty}(g_x^h,g_y^h) = (\mathbf{0},\mathbf{0})$.

By the proof of Proposition 4.7, we see that $\{(x^h,y^h)\}_{h=1}^\infty$ has an accumulation point, say $(x^\star,y^\star)$, and both (4.70) and (4.71) hold. From $f(x^{h_j},y^{h_j}) \leq f_l(x^{h_j},y^{h_j}) + \alpha\epsilon^{h_j} \leq f_l(x^{h_j},x^{h_j}) + \delta^{h_j-1}$, we obtain $f(x^\star,y^\star) \leq \lim_{h\to\infty} f_l(x^h,y^h) + \delta$ by (4.70) and $\delta^k \downarrow \delta$.

Now we show that $(g_x^h,g_y^h) \to (\mathbf{0},\mathbf{0})$ as $h\to\infty$.

By (4.68), $(\hat{x}^h, \hat{y}^h) = (x^h, y^h)$ and Assumption **A5**, we have

$$\nu_h(\|x^h - x^{h+1}\|^2_{\mathcal{M}_{x,h}} + \|y^h - y^{h+1}\|^2_{\mathbf{M}_{y,h}}) = \langle g_x^h, x^h - x^{h+1}\rangle + \langle g_y^h, y^h - y^{h+1}\rangle$$
$$\leq f_{w^{h+1}}(x^h, y^h) - f_{w^{h+1}}(x^{h+1}, y^{h+1}) \leq f_l(x^h, y^h) - f_{w^{h+1}}(x^{h+1}, y^{h+1}).$$

In light of $\nu_h \geq \nu_{min}$, Assumption **A3** and (4.71), we get $\lim_{h\to\infty}(g_x^h, g_y^h) = (\mathbf{0}, \mathbf{0})$. The rest of the proof is identical to that of Proposition 4.7. □

Remark 5.2. *In [86], the condition $\sum_{h=1}^{\infty} \epsilon^h < \infty$ is a requirement for the approximate calculation of proximal points.*

To sum up, the convergence behavior of the Algorithm 4.4, under the inexact setting, can be sharpened by the following theorem.

Theorem 5.1. *Suppose that Assumptions **A3**, **A4** and **A5** hold, $\delta^k \downarrow \delta > 0$, and we choose $\nu_{k+1} \in [\nu_k, \nu_{max}]$ at step 6 of Algorithm 4.4, where $\nu_{max} > 0$ is fixed. Let $\{(\hat{x}^k, \hat{y}^k)\}_{k=1}^{\infty}$ be the sequence of points generated by Algorithm 4.4. There are three possible cases:*

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

2. *If the number of descent steps approaches infinity, then $\overline{\lim}_{k\to\infty} f(\hat{x}^k, \hat{y}^k) \leq \inf f(x, y) + \delta$. In particular, each accumulation point of $\{(\hat{x}^k, \hat{y}^k)\}_{k=1}^{\infty}$ is a δ^{F_1}-optimal solution of (Eigform). If, in addition, $\sum_{k=1}^{\infty} \epsilon^k < \infty$, then either $(\hat{x}^k, \hat{y}^k) \to (x^\star, y^\star) \in arg\min_{x,y} f(x, y)$ or $arg\min_{x,y} f(x, y) = \emptyset$ and $\|(\hat{x}^k, \hat{y}^k)\| \to \infty$ as $k \to \infty$.*

3. *If the number of null steps approaches infinity, then $(\hat{x}^k, \hat{y}^k) \to (\hat{x}, \hat{y})$, which is a δ^{F_1}-optimal solution of (Eigform).*

Proof. Case 1: the number k is finite with $k = K$ on termination. it is exactly the result in Proposition 4.3.

Case 2: the number of descent steps approaches infinity. In this case, without loss of generality, we can assume that for some K, $(\hat{x}^k, \hat{y}^k) = (x^k, y^k)$ for any $k > K$.

Subcase 1: the assumption (4.67) is satisfied. By virtue of Proposition 5.2, the desired results hold.

Subcase 2: the assumption (4.67) is invalid. By the proof of Theorem 4.2, we have $\inf f_l(x^k, y^k) \leq \inf f(x, y)$. Consequently, for any $\epsilon > 0$, there is some integer H such that, for all $k > H$,

$$f(x^k, y^k) - \alpha\epsilon^k \leq f_l(x^k, y^k) < \inf f(x, y) + \epsilon,$$

since $\{f_l(x^k, y^k)\}_{k=1}^{\infty}$ is decreasing. Hence $\overline{\lim}_{k\to\infty} f(x^k, y^k) \leq \inf f(x, y) + \delta$.

If $\sum_{k=1}^{\infty} \epsilon^k < \infty$, then $\lim_{h\to\infty} \epsilon^k = 0$. The rest of the proof is identical to that of Theorem 4.2.

Case 3: a finite number of descent steps.

In this case, there exists iteration N, such that starting with iteration N, the descent test fails in all subsequent iterations in Algorithm 4.4. By Algorithm 4.4, $(\hat{x}^k, \hat{y}^k) = (\hat{x}^N, \hat{y}^N)$ for any $k > N$. Proposition 5.1 shows that $(\hat{x}^k, \hat{y}^k) \to (\hat{x}^N, y^N) =: (\hat{x}, \hat{y})$, which is a δ^{F_1}-optimal solution of (Eigform). $\square$

The following result is the counterpart of Theorem 4.3, and its proof is almost identical to that of Theorem 4.3.

Theorem 5.2. *Let $\mathcal{O}_P = arg\min\{\frac{1}{2}\langle x, \mathcal{Q}(x)\rangle + \langle \hat{s}, x\rangle \mid \mathcal{A}(x) = b, x \succeq \mathbf{0}\}$. Assume that $arg\min_{x,y} f(x, y) \neq \emptyset$, Assumptions* **A1**, **A3**, **A4** *and* **A5** *are satisfied and $\lim_{k\to\infty} \delta^k = \delta > 0$. Let $\{w^k\}$ be the sequence of points generated by Algorithm 4.4. Then the following holds.*

1. *If k is finite with $k = K$ on termination, then $w^{K+1} \in \mathcal{O}_P$.*
2. *If the number of descent steps approaches infinity and $\sum_{k=1}^{\infty} \epsilon^k < \infty$, then all accumulation points of $\{w^k\}_{k=1}^{\infty}$ giving rise to descent steps lie in $\mathcal{O}_P$.*
3. *If the number of null steps approaches infinity, then each accumulation point of $\{w^k\}_{k=1}^{\infty}$ is a δ^{F_1}-optimal solution of (P) in the context of Proposition 5.1.*

In practice, the algorithm will stop at some iteration k due to the given stopping criteria. In this case, we would like to claim that the current stability center $(\hat{x}^k, \hat{y}^k)$ to be an δ^{F1}**-optimal** solution. To this end, we now highlight its sufficient conditions for which the current stability center $(\hat{x}^k, \hat{y}^k)$ is δ^{F1}**-optimal**. The following result is analogous to that in [50, Lemma 2.3].

Proposition 5.3. *i) If $f_l(\hat{x}^k, \hat{y}^k) \leq \inf f_{\widehat{\mathcal{W}}^k}$, then $f(\hat{x}^k, \hat{y}^k) \leq \inf f + \delta$.*

ii) If Assumption **A4** *holds, then $\varpi^k = 0$ implies $f_l(\hat{x}^k, \hat{y}^k) \leq \min f_{\widehat{\mathcal{W}}^k} = f_{\widehat{\mathcal{W}}^k}(\hat{x}^k, \hat{y}^k)$.*

Proof. i) In view of $f_{\widehat{\mathcal{W}}^k}(x, y) \leq f(x, y)$ for all $(x, y) \in \mathbb{V} \times \mathbb{R}^m$, it holds $\inf f_{\widehat{\mathcal{W}}^k} \leq \inf f$.

On the other hand, $f(\hat{x}^k, \hat{y}^k) - \alpha\hat{\epsilon}^k \leq f_l(\hat{x}^k, \hat{y}^k)$. Therefore $f_l(\hat{x}^k, \hat{y}^k) \leq \inf f_{\widehat{\mathcal{W}}^k}$ yields

$$f(\hat{x}^k, \hat{y}^k) \leq \inf f + \alpha\hat{\epsilon}^k \leq \inf f + \delta.$$

ii) By Assumption **A4**, $\eta^k > 0$. Thus $\varpi^k = 0$ results in $(x^{k+1}, y^{k+1}) = (\hat{x}^k, \hat{y}^k)$ and $\tau^k \leq 0$. Consequently, $(g_x^k, g_y^k) = (\mathbf{0}, \mathbf{0})$ and

$$f_l(\hat{x}^k, \hat{y}^k) = \bar{f}^k(\hat{x}^k, \hat{y}^k) + (\tau^k - \alpha\hat{\epsilon}^k) \leq \bar{f}^k(\hat{x}^k, \hat{y}^k) - \alpha\hat{\epsilon}^k \leq \bar{f}^k(\hat{x}^k, \hat{y}^k).$$

Due to (4.14), $\min f_{\widehat{\mathcal{W}}^k} = f_{\widehat{\mathcal{W}}^k}(x^{k+1}, y^{k+1}) = f_{\widehat{\mathcal{W}}^k}(\hat{x}^k, \hat{y}^k)$ since $(g_x^k, g_y^k) = (\mathbf{0}, \mathbf{0})$.

On the other hand, $(g_x^k, g_y^k) = (\mathbf{0}, \mathbf{0})$ results in

$$f_{\widehat{\mathcal{W}}^k}(x^{k+1}, y^{k+1}) \overset{(4.9)}{=} f_{w^{k+1}}(x^{k+1}, y^{k+1}) \overset{(4.15)}{=} \bar{f}^k(x^{k+1}, y^{k+1}) \overset{(4.15)}{=} \bar{f}^k(\hat{x}^k, \hat{y}^k).$$

Therefore, the desired result holds. □

5.2 Updating eigenvalue tolerance δ^k

We see that if the convergence behavior of Algorithm 4.4 is the only concern, one can choose $\{\delta^k\}$ as any arbitrary convergent decreasing sequence. However, the shortcoming of the sequence that decreases too quickly, is pointed out in [70]. It says that such choice would result in wasted work computing precise eigenvalues. Motivated by the following result, in this section we present a choice selection of δ^k.

Theorem 5.3. *Let $\{(x^k, y^k)\}_{k=N}^{\infty}$ be specified by (4.47) and (4.48), respectively. Suppose that $\delta^k \downarrow \delta > 0$ as $k \to \infty$, Assumptions* **A3** *and* **A4** *are satisfied, and we choose $\nu_{k+1} \in [\nu_k, \nu_{max}]$ at step 6 of Algorithm 4.4, where $\nu_{max} > 0$ is fixed. then*

$$\overline{\lim}_{k\to\infty} \nu_k l_d^k \leq \frac{3 - m_L}{1 - m_L}\delta, \tag{5.5}$$

$$\overline{\lim}_{k\to\infty} \eta^k \leq 2\delta. \tag{5.6}$$

Proof. We first prove the second inequality. Using Lemma 5.1, we have, for sufficiently large k,

$$f_l(x^k, y^k) - f_{w^k}(x^k, y^k) \leq \delta. \tag{5.7}$$

By means of

$$\eta^{k-1} \overset{(4.19)}{=} f_u(x^k, y^k) - f_{w^k}(x^k, y^k) \overset{(3.14)}{=} f_l(x^k, y^k) - f_{w^k}(x^k, y^k) + \alpha\epsilon^k,$$

taking into account (5.7) and $\alpha\epsilon^k \leq \delta^{k-1}$, we have (5.6).

By step 6, for $k > N$, it holds $f_l(\hat{x}^k, \hat{y}^k) - f_l(x^{k+1}, y^{k+1}) < m_L\vartheta^k$, whence

$$\begin{aligned} f_l(\hat{x}^k, \hat{y}^k) - f_{w^{k+1}}(x^{k+1}, y^{k+1}) &\overset{(5.1)}{<} \frac{1}{1-m_L}(f_l(x^{k+1}, y^{k+1}) - f_{w^{k+1}}(x^{k+1}, y^{k+1})) \\ &\overset{(4.19)}{\leq} \frac{1}{1-m_L}\eta^k. \end{aligned} \tag{5.8}$$

Furthermore, we have

$$\begin{aligned} \nu_k l_d^k &\overset{(4.38)}{=} \vartheta^k + \alpha\hat{\epsilon}^k - \tau^k \overset{(4.37)}{=} \alpha\hat{\epsilon}^k + f_l(\hat{x}^k, \hat{y}^k) - f_{w^{k+1}}(x^{k+1}, y^{k+1}) - \tau^k \\ &\overset{(4.20)}{\leq} \alpha\hat{\epsilon}^k + f_l(\hat{x}^k, \hat{y}^k) - f_{w^{k+1}}(x^{k+1}, y^{k+1}) \overset{(4.60),(5.8)}{\leq} \delta^{k-1} + \frac{1}{1-m_L}\eta^k. \end{aligned}$$

Therefore, putting together (5.6), we obtain (5.5). □

The above result guides us to decrease δ^k when $\nu_k l_d^k$ or η^k is close to its theoretical limit [70], say within 5% of its theoretical limit. We give the following rule, which is analogous to the one presented in [70].

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

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

5.3 A Lipschitzian error bound for the eigenvalue minimization problem

This section is devoted to a Lipschitzian error bound for (Eigform), followed by an interpretation why the optimality measure (4.25) makes sense in practice. We will see that an upper bound of the optimality measure yields an upper bound estimation of the distance from a given point to the solution set of (Eigform).

Now we turn our attention to spectral inequalities, which hold on simple Euclidean Jordan algebras, see, e.g. [33]. However, we need to establish them to any Euclidean Jordan algebra due to the more general context we are dealing with.

For a Euclidean Jordan algebra $\mathbb{V}$, let

$$\mathcal{J}(\mathbb{V}) := \{c \mid c \text{ is a primitive idempotent in } \mathbb{V}\}.$$

Suppose that $\{c_1, \cdots, c_k\} \subseteq \mathcal{J}(\mathbb{V}),\ 1 \le k \le r$, we denote

$$\mathcal{J}(c_1, \cdots, c_k)^{\perp} := \{d \in \mathcal{J}(\mathbb{V}) \mid d \circ c_i = \mathbf{0},\ i = 1, \cdots, k\}.$$

The next theorem is a generalization to general Euclidean Jordan algebras of the min-max theorem of Hirzebruch, which is known in the setting of simple Euclidean Jordan algebras. The proof of this theorem is based on the following observation.

Lemma 5.2. *Let $\mathbb{V}$ be a Euclidean Jordan algebra with rank $r \ge 2$. For any k primitive idempotents $c_1, \cdots, c_k$, $\mathcal{J}(c_1, \cdots, c_k)^{\perp}$ is nonempty, $1 \le k \le r-1$.*

Proof. By [23, Proposition III.4.4], we can write $\mathbb{V}$ as a direct sum of simple ideals: $\mathbb{V} = \mathbb{V}_1 \oplus \cdots \oplus \mathbb{V}_\kappa$, each $\mathbb{V}_i$ is a simple Euclidean Jordan algebra. Then any primitive idempotent in $\mathbb{V}$ is of the form $(\mathbf{0}, \cdots, \mathbf{0}, c^{(i)}, \mathbf{0}, \cdots, \mathbf{0})$ for some primitive idempotent $c^{(i)} \in \mathbb{V}_i$. Thus, without loss of generality, we may assume that there are a simple Euclidean Jordan algebra $\mathbb{V}_l$ with $1 \le l \le \kappa$ and a subset $\{c_{l_1}, \ldots, c_{l_j}\} \subseteq \{c_1, \cdots, c_k\}$ such that $l_j < rk(\mathbb{V}_l)$. From [43, Lemma 2.4], there is a primitive idempotent $d^{(l)} \in \mathbb{V}_l$ such that $d^{(l)} \circ c_{l_i} = \mathbf{0},\ i = 1, \cdots, j$, whence $\mathcal{J}(c_1, \cdots, c_k)^{\perp}$ is nonempty. $\square$

We are ready to state the main theorem, which is known as the min-max theorem of Hirzebruch [43] for a simple Euclidean Jordan algebra.

Theorem 5.4. *Let $\mathbb{V}$ be a Euclidean Jordan algebra with rank $r \geq 2$. For any $v \in \mathbb{V}$, the following hold:*

$$\lambda_1(v) = \max_{c \in \mathcal{J}(\mathbb{V})} \langle v, c \rangle, \tag{5.9}$$

$$\lambda_r(v) = \min_{c \in \mathcal{J}(\mathbb{V})} \langle v, c \rangle, \tag{5.10}$$

$$\lambda_k(v) = \min_{\{d_1, \cdots, d_{k-1}\} \subseteq \mathcal{J}(\mathbb{V})} \max_{c \in \mathcal{J}(d_1, \cdots, d_{k-1})^{\perp}} \langle v, c \rangle, \;\; k = 2, \cdots, r-1. \tag{5.11}$$

Proof. By Theorem 3.2, we see that $\lambda_1(v) \geq \max_{c \in \mathcal{J}(\mathbb{V})} \langle v, c \rangle$. On the other hand, by applying the spectral decomposition to v, there exists a Jordan frame $\{c_1, \cdots, c_r\}$ such that $v = \sum_{i=1}^r \lambda_i(v) c_i$, whence $c_1 \in \mathcal{J}(\mathbb{V})$ and $\lambda_1(v) = \langle v, c_1 \rangle$ by Corollary 2.1. This proves (5.9). The second equation (5.10) can be obtained by applying this to $-v$.

The proof of (5.11) is almost identical to the proof of Theorem 2.5 in [43]. For $2 \leq k \leq r-1$ and any $\{d_1, \cdots, d_{k-1}\} \subseteq \mathcal{J}(\mathbb{V})$, we set

$$M_1 := \mathcal{J}(d_1, \cdots, d_{k-1})^{\perp} \text{ and } M_2 := \mathcal{J}(c_{k+1}, \cdots, c_r)^{\perp}.$$

Then M_1, M_2, and $M_1 \cap M_2$ are nonempty sets in virtue of Lemma 5.2, hence $\max_{c \in M_1} \langle v, c \rangle > -\infty$. We further denote

$$\gamma_k := \inf_{\{d_1, \cdots, d_{k-1}\} \subseteq \mathcal{J}(\mathbb{V})} \max_{c \in \mathcal{J}(d_1, \cdots, d_{k-1})^{\perp}} \langle v, c \rangle,$$

then

$$\begin{aligned}
\gamma_k &\leq \max_{c \in \mathcal{J}(c_1, \cdots, c_{k-1})^{\perp}} \langle v, c \rangle \\
&= \max_{c \in \mathcal{J}(c_1, \cdots, c_{k-1})^{\perp}} \left\langle \sum_{i=k}^{r} \lambda_i(v) c_i + \lambda_k(v) \sum_{i=1}^{k-1} c_i, c \right\rangle \\
&\leq \max_{c \in \mathcal{J}(\mathbb{V})} \left\langle \sum_{i=k}^{r} \lambda_i(v) c_i + \lambda_k(v) \sum_{i=1}^{k-1} c_i, c \right\rangle = \lambda_k(v),
\end{aligned}$$

where the last equality comes from (5.9).

On the other hand, taking any $d \in M_1 \cap M_2$, we have $\max\limits_{c \in M_1} \langle v, c\rangle \geq \langle v, d\rangle \geq \min\limits_{c \in M_2} \langle v, c\rangle$. Meanwhile, there holds

$$\min_{c \in M_2} \langle v, c\rangle = \min_{c \in M_2} \left\langle \sum_{i=1}^{k} \lambda_i(v) c_i + \lambda_k(v) \sum_{i=k+1}^{r} c_i, c \right\rangle$$
$$\geq \min_{c \in \mathcal{J}(\mathbb{V})} \left\langle \sum_{i=1}^{k} \lambda_i(v) c_i + \lambda_k(v) \sum_{i=k+1}^{r} c_i, c \right\rangle = \lambda_k(v),$$

where the last equality follows from (5.10). Therefore, we can conclude that for any $\{d_1, \cdots, d_{k-1}\} \subseteq \mathcal{J}(\mathbb{V})$, $\max_{c \in \mathcal{J}(d_1, \cdots, d_{k-1})^{\perp}} \geq \lambda_k(v)$, whence $\gamma_k \geq \lambda_k(v)$. Thus

$$\gamma_k = \max_{c \in \mathcal{J}(c_1, \cdots, c_{k-1})^{\perp}} \langle v, c\rangle,$$

which shows that the infimum is obtained. The proof is complete. □

For any $u, v \in \mathbb{V}$, we can generate the Weyl's inequality, which describes the relationship between their eigenvalues and the eigenvalues of $u + v$, to general Euclidean Jordan algebras. Its proof uses the above theorem and is similar to that of the case for simple Euclidean Jordan algebras. We omit the proof here and refer the reader to [72] for details of the simple algebra case.

Theorem 5.5. *(A generalization of Weyl's inequality) Let* $\mathbb{V}$ *be a Euclidean Jordan algebra with rank* $r \geq 2$*. For any* $u, v \in \mathbb{V}$*, it holds*

$$\lambda_i(u) + \lambda_r(v) \leq \lambda_i(u + v) \leq \lambda_i(u) + \lambda_1(v), \ i = 1, \cdots, r. \tag{5.12}$$

For later use, we give a well-known definition of the limiting Fréchet subdifferential, which generalizes the concept of the subdifferential of convex analysis, see, e.g. [57].

Definition 5.2. *Let* $\mathbb{H}$ *be a Hilbert space,* $\psi : \mathbb{H} \to \mathbb{R} \cup \{\infty\}$ *be a lower semi-continuous function and* $x_0 \in \mathbb{H}$ *be such that* $\psi(x_0) < \infty$*. For any* $\epsilon \geq 0$*, the Fréchet* ϵ*-subdifferential is the set*

$$\partial_{\epsilon}^{F} \psi(x) := \left\{ y \in \mathbb{H} \;\middle|\; \liminf_{d \to \mathbf{0}} \frac{\psi(x + d) - \psi(x) - \langle y, d\rangle}{\|d\|} \geq -\epsilon \right\},$$

where $||\cdot||$ *denotes an appropriate norm in* $\mathbb{H}$.
The limiting Fréchet subdifferential of ψ *at* x_0 *is the set*

$$\partial^F \psi(x_0) := \left\{ y \in \mathbb{H} \;\middle|\; \exists\, x_k \to x_0, \epsilon_k \to 0, \text{ and } y_k \to y \text{ with } y_k \in \partial^F_{\epsilon_k} \psi(x_k) \right\}.$$

Remark 5.3. *The paper [76] mentioned that, in a finite dimensional Euclidean space, if, in addition,* ψ *is convex, then* $\partial^F \psi(x_0) = \partial \psi(x_0)$. □

We recall the chain rule for the generalized subdifferentiation of the composition, which comes from Corollary 6.3 in [73]. We use the binary operator $\diamond$ to denote the composition of functions.

Proposition 5.4. *Let* $\mathbb{E}_1,\ \mathbb{E}_2$ *be two Euclidean spaces and* $\Psi : \mathbb{E}_1 \to \mathbb{E}_2$, $\varphi : \mathbb{E}_2 \to \mathbb{R}$. *Suppose that* Ψ *is Lipschitz around* $x_0 \in \mathbb{E}_1$ *and* φ *is Lipschitz around* $\Psi(x_0)$, *then it holds*

$$\partial^F (\varphi \diamond \Psi)(x_0) \subseteq \cup \left\{ \partial^F \langle s, \Psi \rangle (x_0) \;\middle|\; s \in \partial^F \varphi(\Psi(x_0)) \right\},$$

where $\langle s, \Psi \rangle(x) := \langle s, \Psi(x) \rangle$ *and* $\langle \cdot, \cdot \rangle$ *denotes the inner product in* $\mathbb{E}_2$.

We are now ready to present the following fundamental proposition for the study of error bounds, whose proof is very similar to what the authors have done in the proof of Lemma 3 in [47].

Proposition 5.5. *Let* $p : \mathbb{R}^r \to \mathbb{R}$ *be a locally Lipschitz function and* $x \in \mathbb{V}$. *Assume that there exist* $\sigma > 0$ *and* $w \in \mathbb{V}$ *such that* $\lambda_1(\sigma e - w) \sum_{s_i \geq 0} s_i + \lambda_r(\sigma e - w) \sum_{s_i < 0} s_i \leq 0$ *for each* $s \in \partial^F p(\lambda(x))$, *then for each* $v \in \partial^F (p \diamond \lambda)(x)$, *there exists* $\bar{s} \in \partial^F p(\lambda(x))$ *such that*

$$\langle v, w \rangle \geq \sigma \sum_{\imath=1}^{r} \bar{s}_i.$$

Proof. If $w = \mathbf{0}$, then it is trivial.

We now suppose that $w \neq \mathbf{0}$. For each $v \in \partial^F (p \diamond \lambda)(x)$, by Proposition 5.4, there exists $\bar{s} \in \partial^F p(\lambda(x))$ such that $v \in \partial^F \langle \bar{s}, \lambda \rangle (x)$. In light of Definition 5.2, there are $x^k \to x$, $v^k \to v$, $\epsilon_k \to 0$ and $\gamma_k \to 0$ such that

$$\langle \bar{s}, \lambda(x^k + d) - \lambda(x^k) \rangle - \langle v^k, d \rangle + \epsilon_k \|d\|_F \geq 0$$

for any $d \in \mathbb{V}$ with $\|d\|_F \leq \gamma_k$. Let $d = -\frac{\gamma_k}{\|w\|_F} w$ and $q = \sigma e - w$. Then

$$\begin{aligned}
0 \leq & \left\langle \bar{s}, \lambda(x^k - \frac{\gamma_k}{\|w\|_F} w) - \lambda(x^k) \right\rangle + \left\langle v^k, \frac{\gamma_k}{\|w\|_F} w \right\rangle + \epsilon_k \gamma_k \\
= & \left\langle \bar{s}, \lambda(x^k + \frac{\gamma_k}{\|w\|_F} q) - \lambda(x^k) \right\rangle - \frac{\gamma_k}{\|w\|_F} \sigma \sum_{\imath=1}^{r} \bar{s}_i + \left\langle v^k, \frac{\gamma_k}{\|w\|_F} w \right\rangle + \epsilon_k \gamma_k \\
= & \sum_{\bar{s}_i \geq 0} \bar{s}_i \left[\lambda_i(x^k + \frac{\gamma_k}{\|w\|_F} q) - \lambda_i(x^k) \right] + \sum_{\bar{s}_i < 0} \bar{s}_i \left[\lambda_i(x^k + \frac{\gamma_k}{\|w\|_F} q) - \lambda_i(x^k) \right] \\
& - \frac{\gamma_k}{\|w\|_F} \sigma \sum_{\imath=1}^{r} \bar{s}_i + \left\langle v^k, \frac{\gamma_k}{\|w\|_F} w \right\rangle + \epsilon_k \gamma_k.
\end{aligned} \tag{5.13}$$

On the other hand, by virtue of Theorem 5.5, it yields, for any k,

$$\frac{\gamma_k}{\|w\|_F} \lambda_r(q) \leq \lambda_i(x^k + \frac{\gamma_k}{\|w\|_F} q) - \lambda_i(x^k) \leq \frac{\gamma_k}{\|w\|_F} \lambda_1(q), \ i = 1, \cdots, r,$$

therefore (5.13) can be continued with

$$\begin{aligned}
0 \leq & \frac{\gamma_k}{\|w\|_F} \lambda_1(q) \sum_{\bar{s}_i \geq 0} \bar{s}_i + \frac{\gamma_k}{\|w\|_F} \lambda_r(q) \sum_{\bar{s}_i < 0} \bar{s}_i - \frac{\gamma_k}{\|w\|_F} \sigma \sum_{\imath=1}^{r} \bar{s}_i + \left\langle v^k, \frac{\gamma_k}{\|w\|_F} w \right\rangle + \epsilon_k \gamma_k \\
\leq & - \frac{\gamma_k}{\|w\|_F} \sigma \sum_{\imath=1}^{r} \bar{s}_i + \left\langle v^k, \frac{\gamma_k}{\|w\|_F} w \right\rangle + \epsilon_k \gamma_k,
\end{aligned}$$

where the last inequality follows from the assumption. Thus we conclude that

$$\langle v^k, w \rangle + \epsilon_k \|w\|_F \geq \sigma \sum_{\imath=1}^{r} \bar{s}_i,$$

and taking the limit implies $\langle v, w \rangle \geq \sigma \sum_{\imath=1}^{r} \bar{s}_i$. □

Remark 5.4. *1. Let $\sigma > 0$ and $w \in \mathbb{V}$ satisfy the assumptions of the previous proposition, then they should possess the property that*

$$\lambda_1(\sigma e - w) \lambda_r(\sigma e - w) \geq 0.$$

If there are $\sigma > 0$ and $w \in \mathbb{V}$ such that $\lambda_1(\sigma e - w) \leq 0$, then the above proposition is reduced to Lemma 3 in [47] in the framework of symmetric matrices.

2. Special attention is given to the case of $\lambda_1(\sigma e - w) \leq 0$. In this case, we have

$$w \in int(\mathbb{K}) \tag{5.14}$$

and $0 < \sigma \leq \lambda_r(w)$. We name (5.14) a Slater type condition.

Alternatively, the function $f(x,y)$ may be expressed in terms of the composite function, which leads to an alternate description of $\partial f(x,y)$ by Remark 5.3 and Proposition 5.4. More precisely, let $p : \mathbb{R}^r \to \mathbb{R}$ be a locally Lipschitz function, and $\Phi : \mathbb{V} \times \mathbb{R}^m \to \mathbb{V}$ be defined by

$$\Phi(x,y) = \alpha(-\hat{s} - \mathcal{A}^*(y) - \mathcal{Q}(x)) + (\frac{1}{2}\langle x, \mathcal{Q}(x)\rangle + b^T y)e. \tag{5.15}$$

Choosing $p(x_1, \cdots, x_r) = x_1$, we obtain

$$f(x,y) = (p \diamond \lambda \diamond \Phi)(x,y).$$

Plainly, Φ is everywhere Fréchet-differentiable in $\mathbb{V} \times \mathbb{R}^m$. For any given $(x,y) \in \mathbb{V} \times \mathbb{R}^m$, let

$$J\Phi(x,y) : \mathbb{V} \times \mathbb{R}^m \to \mathbb{V},$$

a linear mapping, denote the derivative of Φ at (x,y), and

$$J\Phi(x,y)^* : \mathbb{V} \to \mathbb{V} \times \mathbb{R}^m$$

be the adjoint of $J\Phi(x,y)$.

Via the definition of Fréchet-differentiability, we have, for any $(s,d) \in \mathbb{V} \times \mathbb{R}^m$,

$$J\Phi(x,y)(s,d) = (b^T d + \langle \mathcal{Q}(s), x\rangle)e - \alpha\mathcal{A}^*(d) - \alpha\mathcal{Q}(s),$$

hence, for any $g \in \mathbb{V}$,

$$J\Phi(x,y)^*(g) = (tr(g)\mathcal{Q}(x) - \alpha\mathcal{Q}(g), tr(g)b - \alpha\mathcal{A}(g)).$$

Here is an alternate description of $\partial f(x,y)$.

Theorem 5.6. *Let $\Phi(x,y)$ be specified by (5.15). For a given $(x,y) \in \mathbb{V} \times \mathbb{R}^m$, it holds*

$$\partial f(x,y) \subseteq \left\{ J\Phi(x,y)^*(g) \;\middle|\; g \in \partial(p \diamond \lambda)(\Phi(x,y)) \right\}.$$

Proof. The proof is straightforward by Proposition 5.4. □

At this point, we need to recall one more existing result which plays a key role in our conclusions. We only quote the result in our context.

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

Theorem 5.7. *[103, Theorem 2.2] Suppose that, for some $(x^0,y^0) \in \mathbb{V} \times \mathbb{R}^m$, $0 < \zeta \le +\infty$, $0 < \mu < +\infty$, and $0 < \gamma \le \zeta/(2\mu)$, the set*

$$T := B((x^0,y^0), \gamma/2) \cap \{(x,y) \in \mathbb{V} \times \mathbb{R}^m \mid f(x,y) - f_\star < \gamma\}$$

is nonempty and for all $(g_x, g_y) \in \partial f(x,y)$ and each $(x,y) \in T_1$, where

$$T_1 := B((x^0,y^0), \gamma/2) \cap \{(x,y) \in \mathbb{V} \times \mathbb{R}^m \mid 0 < f(x,y) - f_\star < \gamma\},$$

it holds

$$\|(g_x, g_y)\| \ge \mu^{-1}.$$

Then $\mathcal{O}_\star = arg\min_{x,y} f(x,y)$ is nonempty and

$$dist((x,y), \mathcal{O}_\star) \le \mu(f(x,y) - f_\star) \text{ for all } (x,y) \in T.$$

Moreover, if $(x^0,y^0) \in \mathcal{O}_\star$, then the condition $0 < \gamma \le \zeta/(2\mu)$ can be replaced with $0 < \gamma \le +\infty$.

It is known that (5.14) is a Slater type condition. When applied to $J\Phi(x,y)(s,d)$, it gives the existence of some $(s,d) \in \mathbb{V} \times \mathbb{R}^m$ such that $J\Phi(x,y)(s,d)$ lies in $int(\mathbb{K})$ for all $(x,y) \in \mathbb{V} \times \mathbb{R}^m$. In the framework of symmetric matrices, a global error bound for the special case when $\Phi(x,y) = -\hat{s} - \mathcal{A}^*(y)$ under this type condition is given in [47, Corollary 2]. Using similar arguments, we can derive the following Lipschitzian error bound for (Eigform), which is the more general context we are dealing with.

Theorem 5.8. *Suppose that Assumption* **A1** *is satisfied. Let* $\{(x^k, y^k)\}_{k=1}^N$ *be a finite sequence of points. Further, suppose that there exist* $\sigma > 0$ *and* $(s, d) \in \mathbb{V} \times \mathbb{R}^m$ *such that*

$$\lambda_1(\sigma e - J\Phi(x^k, y^k)(s, d)) \leq 0 \;\; \textit{for any} \;\; k = 1, \cdots, N, \tag{5.16}$$

where $\Phi(x, y)$ *is specified by (5.15). Then* $\mathcal{O}_\star$ *is nonempty and*

$$\textit{dist}((x^k, y^k), \mathcal{O}_\star) \leq \frac{\sqrt{\|s\|_F^2 + \|d\|_2^2}}{\sigma}(f(x^k, y^k) - f_\star) \;\; \textit{for any} \;\; k = 1, \cdots, N. \tag{5.17}$$

Proof. In light of Proposition 3.1, we see that

$$\mathcal{O}_D = \arg\min \left\{ g(x, y) \;\middle|\; \mathcal{Q}(x) + \mathcal{A}^*(y) + \hat{s} \in \mathbb{K} \right\},$$
$$\mathcal{O}_\star = \left\{ (x^\star, y^\star) + \tau(0, \bar{y}) \;\middle|\; (x^\star, y^\star) \in \mathcal{O}_D, \tau \in \mathbb{R} \right\}.$$

Plainly, Assumption **A1** and $f_\star > -\infty$ imply that $\mathcal{O}_D$ is nonempty, whence $\mathcal{O}_\star$ is also nonempty. Then one can choose the desired point in Theorem 5.7 such that $(x^0, y^0) \in \mathcal{O}_\star$, whence $\gamma = \infty$ by Theorem 5.7 so that $\{(x^k, y^k)\}_{k=1}^N \subseteq T$.

By setting $p(x_1, \cdots, x_r) = x_1 - f_\star$, we have $f(x, y) - f_\star = (p \diamond \lambda \diamond \Phi)(x, y)$. It is easy to see that $\partial p(\lambda(\Phi(x, y))) = \{(1, 0, \cdots, 0)\}$ for all $(x, y) \in \mathbb{V} \times \mathbb{R}^m$. 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 5.6, there is $g^k \in \partial(p \diamond \lambda)(\Phi(x^k, y^k))$ such that $(g_{x^k}, g_{y^k}) = J\Phi(x^k, y^k)^*(g^k)$. Using (5.16), by Proposition 5.5, we have

$$\langle (g_{x^k}, g_{y^k}), (s, d) \rangle = \langle J\Phi(x^k, y^k)^*(g^k), (s, d) \rangle = \langle g^k, J\Phi(x^k, y^k)(s, d) \rangle \geq \sigma,$$

hence $\|(g_{x^k}, g_{y^k})\| := \sqrt{||g_{x^k}||_F^2 + ||g_{y^k}||_2^2} \geq \frac{\sigma}{\sqrt{\|s\|_F^2 + \|d\|_2^2}}$ and the proof is complete. □

We now return to the construction of the optimality measure ϖ^k. Suppose that the conditions in Theorem 5.8 hold and at iteration $k+1$ it satisfies $\varpi^{k+1} \leq \varepsilon$. Putting together (4.23) and (4.24), via Theorem 5.8 and the fact that $\{(\hat{x}^k, \hat{y}^k)\}$ is a subsequence of $\{(x^k, y^k)\}$, we obtain that one of the following inequalities is satisfied

$$\textit{dist}((\hat{x}^k, \hat{y}^k), \mathcal{O}_\star) \leq \frac{\varepsilon}{\frac{\sigma}{\sqrt{\|s\|_F^2 + \|d\|_2^2}} - \varepsilon} \tag{5.18}$$

$$dist((x^{k+1},y^{k+1}),\mathcal{O}_\star) \le \frac{\varepsilon}{\frac{\sigma}{\sqrt{\|s\|_F^2+\|d\|_2^2}}-\varepsilon}. \tag{5.19}$$

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

Example 5.1. *We continue with Example 3.1. We show that Problem (3.5) satisfies condition (5.16) for the special case where $\mathcal{Q}$ has the property that there is some scalar ρ with $\rho < \|\mathcal{Q}\|_2$ such that $tr(\mathcal{Q}(X)) = \rho tr(X)$, for instance $\mathcal{Q}(X) = H \cdot X$ with all diagonal entries being the same and $\mathcal{Q}(X) = UXU^T$ with U being an orthogonal matrix. In practice, we will terminate Algorithm 4.4 at iteration $N \ge 1$ with $\varpi^N \le \varepsilon$. We will show that the conditions in Theorem 5.8 hold, whence the sequence of points $\{(X^k,y^k)\}_{k=1}^N$ generated by Algorithm 4.4 has one of the properties (5.18) and (5.19).*

Indeed, using $\mathcal{M}_{X,k} = \mathcal{I} - \frac{1}{\nu_k}\mathcal{Q}$, where $\nu_k > \|\mathcal{Q}\|_2$ for $k = 0,1,\cdots,$ together with (4.10), we have

$$\nu_{min}, \nu_{k-1} > \|\mathcal{Q}\|_2 > \rho \textit{ and } X^k = \widehat{X}^{k-1} + \frac{1}{\nu_{k-1}}\mathcal{Q}(W^k - \widehat{X}^{k-1}) \textit{ for all } k = 1,2,\cdots,N.$$

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

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

since $\widehat{X}^k$ is one of $X^0,\cdots,X^k$.

By setting

$$\tau = \max\{tr(\widehat{X}^0),\cdots,tr(\widehat{X}^{N-1})\} < n,$$

we have

$$\begin{aligned} tr(X^k) - n =& \frac{\rho}{\nu_{k-1}}tr(W^k) + (1-\frac{\rho}{\nu_{k-1}})tr(\widehat{X}^{k-1}) - n \\ \le& \frac{n\rho}{\nu_{k-1}} + (1-\frac{\rho}{\nu_{k-1}})\tau - n \\ =& (n-\tau)(\frac{\rho}{\nu_{k-1}} - 1) \end{aligned}$$

$$\leq (n-\tau)(\frac{\rho}{\nu_{min}}-1) < 0$$

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

On the other hand, it holds $\sigma I - J\Phi(X^k, y^k)(D, d) = (\sigma - \mathbf{1}^T d - \langle X^k, \mathcal{Q}(D)\rangle)I + n(\mathcal{A}^*(d) + \mathcal{Q}(D))$. *Therefore, we can choose* $(\mathcal{Q}(D), d) = (\iota I, \mathbf{0})$ *with* $\iota < 0$ *such that*

$$0 < \sigma \leq \iota(n-\tau)(\frac{\rho}{\nu_{min}}-1),$$

then

$$\sigma I - J\Phi(X^k, y^k)(D, d) = \sigma I + \iota(n - tr(X^k))I \preceq \mathbf{0} \text{ for all } k = 1, \cdots, N.$$

Choosing $\sigma = \iota(n-\tau)(\frac{\rho}{\nu_{min}}-1)$, *by Theorem 5.8, (4.23) and (4.24), we have one of the following inequalities*

$$dist((\hat{x}^k, \hat{y}^k), \mathcal{O}_\star) \leq \frac{\varepsilon}{\frac{\iota(n-\tau)}{\|D\|_F}(\frac{\rho}{\nu_{min}}-1)-\varepsilon}$$
$$dist((x^{k+1}, y^{k+1}), \mathcal{O}_\star) \leq \frac{\varepsilon}{\frac{\iota(n-\tau)}{\|D\|_F}(\frac{\rho}{\nu_{min}}-1)-\varepsilon}.$$

Note that $\iota n = tr(\mathcal{Q}(D)) = \rho tr(D)$. *In particular, if there is a matrix* D, *obtained when an identity matrix is multiplied by a scalar, such that* $\mathcal{Q}(D) = \iota I$. *Then the above two inequalities reduce to*

$$dist((\hat{x}^k, \hat{y}^k), \mathcal{O}_\star) \leq \frac{\varepsilon}{\frac{\rho(n-\tau)}{\sqrt{n}}(1-\frac{\rho}{\nu_{min}})-\varepsilon}$$
$$dist((x^{k+1}, y^{k+1}), \mathcal{O}_\star) \leq \frac{\varepsilon}{\frac{\rho(n-\tau)}{\sqrt{n}}(1-\frac{\rho}{\nu_{min}})-\varepsilon}.$$

Chapter 6

A Lipschitzian error bound for convex quadratic symmetric cone programming

In this chapter, we are interested in the following symmetric cone linear complementarity problem (SCLCP), which is a generalization of semidefinite linear complementarity problems stated in [54]: given the closure $\mathbb{K}$ of the symmetric cone associated with a Euclidean Jordan algebra $\mathbb{V}$, find an $(x, z) \in \mathbb{V} \times \mathbb{V}$ such that

$$\text{(SCLCP)} \qquad (x, z) \in \mathcal{G}, x \in \mathbb{K}, z \in \mathbb{K}, \text{ and } \langle x, z \rangle = 0.$$

Here $\mathcal{G}$ is an affine subspace of $\mathbb{V} \times \mathbb{V}$ with $dim(\mathcal{G}) = dim(\mathbb{V})$, say equal to n. Let $\mathcal{F}$, $\mathcal{F}^\circ$ and $\mathcal{F}^\star$ denote, respectively, the feasible solution set $\{(x, z) \in \mathcal{G} \mid x \in \mathbb{K}, z \in \mathbb{K}\}$, the strictly feasible solution set $\{(x, z) \in \mathcal{G} \mid x \in int(\mathbb{K}), z \in int(\mathbb{K})\}$, and the solution set $\{(x, z) \in \mathcal{F} \mid \langle x, z \rangle = 0\}$ of (SCLCP).

We focus on the monotone (SCLCP), which serves a pair of primal-dual of convex quadratic symmetric cone programs (P) and (D) as a special case. Under Assumptions **A1** and **A2**, we can state a common necessary and sufficient optimality condition for x to be a minimum solution of (P) and (x, y, z) to be a minimum solution of (D) as follows

$$\mathcal{A}(x) = b, \ z - \mathcal{Q}(x) - \mathcal{A}^*(y) = \hat{s}, \ \langle x, z \rangle = 0, \ x, z \in \mathbb{K}. \tag{6.1}$$

Let

$$\mathcal{G} = \left\{ (x,z) \in \mathbb{V} \times \mathbb{V} \;\middle|\; \mathcal{A}(x) = b,\; z - \mathcal{Q}(x) - \mathcal{A}^*(y) = \hat{s} \text{ for some } y \in \mathbb{R}^m \right\},$$

then (6.1) implies that (x, y, z) is a primal-dual optimal solution for (P) and (D) if and only if $(x, z) \in \mathcal{F}^\star$. By Proposition 2.1 in [24], it holds $dim(\mathcal{G}) = dim(\mathbb{V})$ (see paragraph immediately after Proposition 2.1 of [24]). In addition, the affine subspace $\mathcal{G}$ enjoys monotonicity. In effect, for every (x, z), $(x', z') \in \mathcal{G}$, we have

$$\langle x - x', z - z' \rangle = \langle x - x', \mathcal{Q}(x - x') + \mathcal{A}^*(y - y') \rangle = \langle x - x', \mathcal{Q}(x - x') \rangle \geq 0.$$

Therefore we see that the monotone (SCLCP) is at least as general as the primal-dual pair of the convex quadratic symmetric cone problems. On the other hand, the monotone (SCLCP) can be reformulated as a convex quadratic cone program: $\min\{\langle x, z \rangle \mid (x, z) \in \mathcal{G}, x \in \mathbb{K}, z \in \mathbb{K}\}$.

We should mention that we need Assumption **A2** to ensure that the set of strictly feasible solutions for (P) is nonempty. However, in practice, choice of a strictly feasible solution is not an easy task even in linear programming. It is one of research topics treated with fervor, see, e.g., [9, 65, 96].

In this chapter, we consider a sequence of strictly feasible solutions within a wide neighborhood of central trajectory for the monotone (SCLCP). Under assumptions of strict complementarity and Slater's condition, we establish this sequence leads to a strictly complementary solution, and provide four necessary and sufficient conditions of a Lipschitizan error bound for the monotone (SCLCP). Consequently, we obtain four different ways of depicting a sequence of primal-dual solutions to converge linearly with their duality gaps, which lies within a wide neighborhood of the central path of the primal-dual pair (P) and (D).

6.1 Some properties of the solution set of the monotone symmetric cone linear complementarity problem

In the setting of symmetric matrices of order n, it was shown in [54] that, under Slater's condition, there is a unitary similarity transformation of the matrices in the solution set of the monotone (SCLCP) such that all solutions (X, Z) are of the form

$$X = \begin{bmatrix} X_{BB} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \; Z = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Z_{NN} \end{bmatrix}, X_{BB}, Z_{NN} \succeq \mathbf{0},$$

where $B \cup N = \{1, \cdots, n\}$ and $B \cap N = \emptyset$. The aim of this section is to extend this result to the context of a Euclidean Jordan algebra. Under the additional assumption of strict complementarity, we further obtain that a sequence of strictly feasible solutions of the monotone (SCLCP) converges to a strictly complementary solution.

In the sequel, we impose the following assumptions on (SCLCP).

- **(C1)** The $n-$dimensional affine subspace $\mathcal{G}$ associated with (SCLCP) is monotone, i.e., $\langle x - x', z - z' \rangle \geq 0$ for every $(x, z), (x', z') \in \mathcal{G}$.
- **(C2)** The strictly feasible solution set $\mathcal{F}^\circ$ is nonempty.
- **(C3)** The (SCLCP) has at least one solution $(x^\star, z^\star) \in \mathcal{F}^\star$ which is *strictly complementarity*, that is a solution with $x^\star + z^\star \in int(\mathbb{K})$.
- **(C4)** For the sequence $\{(x(l), z(l))\}_{l=1}^\infty \subseteq \mathcal{F}^\circ$, it holds

$$(\rho :=) \inf_{l=1,2,\ldots} \{\lambda_r(P_{x(l)^{1/2}}(z(l)))/\langle x(l), z(l)\rangle\} > 0.$$

Assumption **(C4)** is equivalent to keeping the iterates in the wide neighborhood of the central trajectory

$$\{(x, z) \in \mathcal{F}^\circ \mid \lambda_r(P_{x^{1/2}}(z)) \geq \rho\langle x, z\rangle\}.$$

This is the context in which we want to deal with.

We now make some comments on Assumption **(C2)**.

In light of Assumption **(C2)**, the solution set $\mathcal{F}^\star$ possesses the following properties that will be used later on. Such similar properties were given by Theorem 6.1 of [54] in the context of semidefinite linear complementarity problems.

Lemma 6.1. *Suppose that the monotone (SCLCP) satisfies Assumption* **(C2)**.

(1) The solution set $\mathcal{F}^\star$ is a nonempty closed convex set.

(2) There exist disjoint subsets B, N of the index set $\{1, \cdots, r\}$ with $B = \{1, \cdots, k\}$ such that for every $(x, z) \in \mathcal{F}^\star$,

$$\begin{aligned} &B \cup N = \{1, \cdots, r\}, \\ &x \in \left\{ x \in \mathbb{K} \;\middle|\; x_{BB} \in \mathbb{K}^{(k)}, x_{BN} = \mathbf{0}, x_{NN} = \mathbf{0} \right\}, \\ &z \in \left\{ z \in \mathbb{K} \;\middle|\; z_{NN} \in \mathbb{K}^{(r-k)}, z_{BN} = \mathbf{0}, z_{BB} = \mathbf{0} \right\}, \end{aligned}$$

where x_{BB}, x_{BN}, x_{NN} are specified by (2.4).

Proof. (1) Clearly, the solution set $\mathcal{F}^\star$ is closed. The nonemptiness of the solution set follows from Corollary 2.3 in [24]. For the convexity of the solution set, see proof of 1 in [54, Theorem 6.1].

(2) Let d be the dimension of the affine subspace spanned by the solution set $\mathcal{F}^\star$. It follows that there exist $d+1$ pairs of elements $(x^0, z^0), (x^1, z^1), \cdots, (x^d, z^d) \in \mathcal{F}^\star$, such that for any $(x, z) \in \mathcal{F}^\star$,

$$(x, z) = \sum_{j=0}^{d} \alpha_j (x^j, z^j) \text{ with } \sum_{j=0}^{d} \alpha_j = 1.$$

Set $(\bar{x}, \bar{z}) = \frac{1}{d+1} \sum_{j=0}^{d} (x^j, z^j)$, then $(\bar{x}, \bar{z}) \in \mathcal{F}^\star$ by the convexity of $\mathcal{F}^\star$. We further denote by $S = \{v \in \mathbb{V} \mid \langle v, L_{\bar{x}}(v) \rangle = 0\}$.

Let $\bar{x} = \sum_{i=1}^{r} \lambda_i(\bar{x}) c_i$ be a spectral decomposition of $\bar{x}$, and

$$B = \{i \mid \lambda_i(\bar{x}) > 0\} =: \{1, \cdots, k\}, N = \{i \mid \lambda_i(\bar{x}) = 0\} =: \{k+1, \cdots, r\}.$$

We shall show that for every $(x, z) \in \mathcal{F}^\star$, $x \circ c_i = \mathbf{0}, i \in N$ and $z \circ c_i = \mathbf{0}, i \in B$, whence $x \in \mathbb{V}^{(k)}$ and $z \in \mathbb{V}^{(r-k)}$ by Proposition 2.8 and Proposition 2.10. Thus the desired results hold in light of $x, z \in \mathbb{K}$ and (iii) of Proposition 2.9.

For any $v \in S$, we have

$$0 = \langle v, L_{\bar{x}}(v) \rangle = \frac{1}{d+1} \sum_{j=0}^{d} \langle v^2, x^j \rangle,$$

whence $\langle v^2, x^j \rangle = 0, j = 0, 1, \cdots, d$, since $v^2, x^j \in \mathbb{K}$.

On the other hand, for $i \in N$, we have $\langle \bar{x}, c_i \rangle = \lambda_i(\bar{x}) = 0$, whence $\langle c_i, L_{\bar{x}}(c_i) \rangle = 0$, and then $c_i \in S, i \in N$, thus $\langle c_i, x^j \rangle = 0, i \in N, j = 0, 1, \cdots, d$. Since any x with $(x, z) \in \mathcal{F}^\star$ is an affine combination of $x^0, x^1, \cdots, x^d$, we conclude that $\langle x, c_i \rangle = 0, i \in N$ for any $(x, z) \in \mathcal{F}^\star$.

Noting that $(\bar{x}, \bar{z}) \in \mathcal{F}^\star$ implies $\bar{x} \circ \bar{z} = \mathbf{0}$, we have for $i \in B$,

$$0 = \langle c_i, \bar{x} \circ \bar{z} \rangle = \langle c_i \circ \bar{x}, \bar{z} \rangle = \lambda_i(\bar{x}) \langle c_i, \bar{z} \rangle,$$

whence $\langle c_i, \bar{z} \rangle = 0, i \in B$. Consequently, $\langle c_i, z^j \rangle = 0, i \in B, j = 0, 1, \cdots, d$. Therefore, $\langle z, c_i \rangle = 0, i \in B$ for any $(x, z) \in \mathcal{F}^\star$. □

For notational convenience, we use $\sigma(l)$ to denote $\langle x(l), z(l) \rangle$ and define

$$W(l) = \sigma(l)^{-2} P_{x(l)} P_{z(l)}, \tag{6.2}$$

$$u(l) = \sigma(l)^{-1} P_{x(l)^{1/2}}(z(l)). \tag{6.3}$$

In addition, if $\mathbb{V}$ is special, we specify

$$v(l) = \sigma(l)^{-1} x(l) z(l). \tag{6.4}$$

By restricting $\mathbb{V} = \mathbb{S}^n$ and $\mathbb{K} = \mathbb{S}^n_+$ in the general formulation of convex quadratic symmetric cone programming in (P) and (D), we obtain a primal-dual pair of quadratic semidefinite programs. When there is no quadratic term in the primal-dual pair (P) and (D), we have the standard semidefinite programming (SDP). In this case, two necessary and sufficient conditions of a Lipschitzian error bound for SDP were well-studied in [19] under the assumptions of strict complementarity and Slater's condition. We will derive similar results on the monotone (SCLCP).

In what follows, we use $(x^\star, z^\star) \in \mathcal{F}^\star$ to denote a fixed strictly complementary solution. Let index sets B and N be given by Lemma 6.1. We shall, with no loss of generality, make the following assumption for (SCLCP).

(**A6**) We can partition $x^\star$ and $z^\star$ as

$$x^\star = x^\star_{BB} \in int(\mathbb{K}^{(k)}), \tag{6.5}$$

and

$$z^\star = z^\star_{NN} \in int(\mathbb{K}^{(r-k)}), \tag{6.6}$$

respectively, where $B = \{1, \cdots, k\}$ and $N = \{k+1, \cdots, r\}$ with $0 \le k \le r$.

This gives rise to a similar partitioning of all $x \in \mathbb{V}$,

$$x = x_{BB} + x_{BN} + x_{NN},$$

where $x_{BB} \in \mathbb{V}_{BB}, x_{BN} \in \mathbb{V}_{BN}, x_{NN} \in \mathbb{V}_{NN}$, and $\mathbb{V}_{BB}, \mathbb{V}_{BN}, \mathbb{V}_{NN}$ are specified by (2.2). In addition, if $\mathbb{V}$ is special and derived from the associative algebra $\mathbb{A}$, the above partition gives rises to the associative Peirce decomposition of all elements $a \in \mathbb{A}$,

$$a = a_{BB} + a_{BN} + a_{NB} + a_{NN},$$

where $a_{BB} \in \mathbb{A}_{11}, a_{BN} \in \mathbb{A}_{12}, a_{NB} \in \mathbb{A}_{21}, a_{NN} \in \mathbb{A}_{22}$. In particular, if the associative algebra $\mathbb{A}$ has an involution $*$, then for any $a \in \mathbb{V}$,

$$a = a_{BB} + a_{BN} + a^*_{BN} + a_{NN},$$

where $a_{BB} \in \mathbb{V}_{BB}, a_{BN} + a^*_{BN} \in \mathbb{V}_{BN}, a_{NN} \in \mathbb{V}_{NN}$.

First of all, we consider a sequence of strictly feasible solutions of the monotone (SCLCP), and prove that any limit point of this sequence is a strictly complementary solution under Assumption **A6**. Such limiting behavior of the sequence will be connected with the error bound results for the monotone (SCLCP). We will need the following simple bounds on the inner product of two elements of $\mathbb{K}$.

Lemma 6.2. *For $x, y \in \mathbb{K}$ there holds*

$$\lambda_r(x)||y||_F \le \lambda_r(x)tr(y) \le \langle x, y\rangle \le \lambda_1(x)tr(y).$$

Proof. The first inequality follows immediately from the almost trivial inequality

$$||y||_F \leq tr(y) \text{ for all } y \in \mathbb{K}.$$

Let $x = \lambda_1(x)c_1 + \cdots + \lambda_r(x)c_r$ be the spectral decomposition of x, then

$$\begin{aligned}\langle x, y\rangle &= \langle \lambda_1(x)c_1 + \cdots + \lambda_r(x)c_r, y\rangle \\ &\geq \lambda_r(x)\langle c_1 + \cdots + c_r, y\rangle = \lambda_r(x)tr(y).\end{aligned}$$

Similarly, we have $\langle x, y\rangle \leq \lambda_1(x)tr(y)$. □

Lemma 6.3. *Assume* **(C1)**, **(C2)**, **(C3)**, **(C4)**, *Assumption* **A6**, *and that the sequence of strictly feasible solutions* $\{(x(l), z(l))\}_{l=1}^{\infty}$ *has* $\lim_{l\to\infty} \sigma(l) = 0$. *If* $(x(\infty), z(\infty))$ *is a limit point of the sequence, then*

$$\begin{aligned}x(\infty) &\in \{x \in \mathbb{K} \mid x_{BB} \in int(\mathbb{K}^{(k)}), x_{BN} = \mathbf{0}, x_{NN} = \mathbf{0}\}, \\ z(\infty) &\in \{z \in \mathbb{K} \mid z_{NN} \in int(\mathbb{K}^{(r-k)}), z_{BN} = \mathbf{0}, z_{BB} = \mathbf{0}\}.\end{aligned}$$

Proof. The assumption $\lim_{l\to\infty} \sigma(l) = 0$ implies that $\langle x(\infty), z(\infty)\rangle = 0$, from which it follows that $(x(\infty), z(\infty)) \in \mathcal{F}^\star$, whence

$$\begin{aligned}x(\infty)_{BB} \in \mathbb{K}^{(k)}, x(\infty)_{BN} = \mathbf{0}, x(\infty)_{NN} = \mathbf{0}, \\ z(\infty)_{NN} \in \mathbb{K}^{(r-k)}, z(\infty)_{BN} = \mathbf{0}, z(\infty)_{BB} = \mathbf{0},\end{aligned}$$

by (2) of Lemma 6.1. Therefore, it suffices to show that

$$rk(x(\infty)) \geq k \text{ and } rk(z(\infty)) \geq r - k.$$

We shall prove the former and remark that the later can be similarly proven.

For any $l = 1, 2, \cdots$, $x(l) \in int(\mathbb{K})$, we denote the respective spectral decomposition of these elements by

$$x(l) = \sum_{i=1}^{r} \lambda_i(x(l))c_i(x(l)).$$

Let $\{x(l_j)\}_{j=1}^{\infty}$ be a subsequence converging to $x(\infty)$. For each $i = 1, \cdots, r$, $\lambda_i(\cdot)$ is continuous. Consequently, these sequences $\{\lambda_i(x(l_j))\}_{j=1}^{\infty}$ are bounded. The sequences $\{c_i(x(l_j))\}_{j=1}^{\infty}$ are bounded by the compactness of the set of primitive idempotents. Passing to a subsequence if necessary, we may assume that

$$\lim_{j\to\infty} \lambda_i(x(l_j)) = \lambda_i(\infty), i = 1, \cdots, r,$$

and $\{c_i(x(l_j))\}_{j=1}^{\infty}$ converges to $c_i(\infty)$ such that

$$x(\infty) = \sum_{i=1}^{r} \lambda_i(\infty) c_i(\infty).$$

In view of the continuity of eigenvalue functions and (iii) of Corollary 2.1, it is seen that $\{c_1(\infty), \cdots, c_r(\infty)\}$ is a Jordan frame.

By Assumption **(C4)**, we have $\lambda_r(u(l)) \geq \rho > 0$. In view of assumptions **(C1)** and **(C3)**, it yields $\langle x(l_j) - x^\star, z(l_j) - z^\star \rangle \geq 0$, whence

$$\begin{aligned}
\sigma(l_j) &= \langle x(l_j), z(l_j) \rangle + \langle x^\star, z^\star \rangle \geq \langle x(l_j), z^\star \rangle + \langle z(l_j), x^\star \rangle \\
&\geq \langle z(l_j), x^\star \rangle \overset{(6.3)}{=} \sigma(l_j) \langle P_{x(l_j)^{-1/2}}(u(l_j)), x^\star \rangle \\
&= \sigma(l_j) \langle u(l_j), P_{x(l_j)^{-1/2}}(x^\star) \rangle \geq \rho\sigma(l_j) tr(P_{x(l_j)^{-1/2}}(x^\star)) \\
&= \rho\sigma(l_j) tr(x(l_j)^{-1} \circ x^\star) = \rho\sigma(l_j) \sum_{i=1}^{r} \frac{1}{\lambda_i(x(l_j))} \langle c_i(x(l_j)), x^\star \rangle \\
&\geq \frac{\rho\sigma(l_j)}{\lambda_i(x(l_j))} \langle c_i(x(l_j)), x^\star \rangle,
\end{aligned}$$

where the penultimate inequality follows (iv) of Proposition 2.2, Lemma 6.2 and $\lambda_r(u(l_j)) \geq \rho$. This means

$$\langle c_i(x(l_j)), x^\star \rangle \leq \frac{\lambda_i(x(l_j))}{\rho}, i = 1, \cdots, r.$$

Letting j approach infinity, we achieve

$$\langle c_i(\infty), x^\star \rangle = 0 \text{ whenever } \lambda_i(\infty) = 0,$$

whence

$$c_i(\infty) \circ x^\star = \mathbf{0} \text{ if } \lambda_i(\infty) = 0. \tag{6.7}$$

We denote by $I_1 = \{i \mid \lambda_i(\infty) > 0\}$ and $I_2 = \{i \mid \lambda_i(\infty) = 0\}$. Let

$$x^\star = \sum_{i=1}^{r} x_i^\star c_i(\infty) + \sum_{i<j} x_{ij}^\star$$

be the Peirce decomposition of $x^\star$ with respect to the Jordan frame $\{c_1(\infty), \cdots, c_r(\infty)\}$. By (6.7), we have

$$x^\star = \sum_{i \in I_1} x_i^\star c_i(\infty) + \sum_{i<j, i,j \in I_1} x_{ij}^\star,$$

whence $x^\star \in \mathbb{V}^{(|I_1|)}$ by (ii) of Proposition 2.9. Since

$$rk(\mathbb{V}^{(|I_1|)}) = |I_1|$$

by (i) of Proposition 2.9, we conclude that

$$rk(x(\infty)) = |I_1| \geq rk(x^\star) = k,$$

where the last equality is by (6.5). □

6.2 Necessary and sufficient conditions for a Lipschitzian error bound

A central trajectory following method for the monotone (SCLCP) is of theoretical importance, see, for instance, [24,59] and the references therein. It generates a sequence of strictly feasible solutions $\{(x(l), z(l))\}_{l=1}^{\infty}$ in a certain neighborhood. This section is devoted to the following Lipschitzian error bound for the monotone (SCLCP) under the assumption of strict complementarity (see **(C3)**), which states the convergence rate of such iterates,

$$\text{(LB)} \qquad dist((x(l), z(l)), \mathcal{F}^\star) = O(\langle x(l), z(l) \rangle),$$

where $dist((x,z),\mathcal{S})$ denotes the value $\inf\left\{\sqrt{||x-u||_F^2+||z-v||_F^2}\;\middle|\;(u,v)\in\mathcal{S}\right\}$ for each $(x,z)\in\mathbb{V}\times\mathbb{V}$ and each $\mathcal{S}\subseteq\mathbb{V}\times\mathbb{V}$. We will derive some necessary and sufficient conditions for such Lipschitzian error bound.

First of all, We study the asymptotic behavior of the "block" forms $x(l)_{BB}, x(l)_{BN}$, $x(l)_{NN}$ of

$$x(l) = x(l)_{BB} + x(l)_{BN} + x(l)_{NN},$$

and those of $z(l)$ as $l\to\infty$.

Lemma 6.4. *Assume* **(C1)**, **(C2)**, **(C3)**, *and the sequence* $\{(x(l),z(l))\}_{l=1}^{\infty}\subseteq\mathcal{F}^\circ$, *then*

$$||x(l)_{BB}||_F = O(1) \text{ and } ||z(l)_{NN}||_F = O(1), \tag{6.8}$$

$$||x(l)_{NN}||_F = O(\sigma(l)) \text{ and } ||z(l)_{BB}||_F = O(\sigma(l)), \tag{6.9}$$

$$||x(l)_{BN}||_F = O(\sqrt{\sigma(l)}) \text{ and } ||z(l)_{BN}||_F = O(\sqrt{\sigma(l)}). \tag{6.10}$$

Proof. We can pick some $(x^\circ,z^\circ)\in\mathcal{F}^\circ$ by Assumption **(C2)**. For every $l\geq 1$, there holds

$$\langle x(l)-x^\circ, z(l)-z^\circ\rangle \geq 0$$

by Assumption **(C1)**. Using this property, together with Lemma 6.2, we get

$$\lambda_r(z^\circ)||x(l)||_F + \lambda_r(x^\circ)||z(l)||_F \leq \langle x(l),z(l)\rangle + \langle x^\circ,z^\circ\rangle = \sigma(l) + \langle x^\circ,z^\circ\rangle.$$

Since $x^\circ, z^\circ\in int(\mathbb{K})$, we obtain

$$||x(l)||_F = O(1+\sigma(l)) \text{ and } ||z(l)||_F = O(1+\sigma(l)). \tag{6.11}$$

By virtue of $\langle x(l)-x^\star, z(l)-z^\star\rangle\geq 0$, we have

$$\begin{aligned}\sigma(l) &= \langle x(l),z(l)\rangle + \langle x^\star,z^\star\rangle \geq \langle x(l),z^\star\rangle + \langle z(l),x^\star\rangle\\ &\overset{(6.5),(6.6)}{=} \langle x(l)_{NN}, z^\star_{NN}\rangle + \langle z(l)_{BB}, x^\star_{BB}\rangle\\ &\overset{Lemma\ 6.2}{\geq} \lambda_{min}(z^\star_{NN},\mathbb{V}_{NN})||x(l)_{NN}||_F + \lambda_{min}(x^\star_{BB},\mathbb{V}_{BB})||z(l)_{BB}||_F.\end{aligned}$$

Since $\lambda_{min}(x^\star_{BB}, \mathbb{V}_{BB}) > 0$ and $\lambda_{min}(z^\star_{NN}, \mathbb{V}_{NN}) > 0$, we have (6.9). Taking into account that

$$||x(l)_{BB}||_F + ||x(l)_{NN}||_F \leq \sqrt{3}||x(l)||_F,$$

putting together (6.11) and (6.9), we obtain (6.8).

We shall prove the first equation of (6.10), the second one can be proved similarly. As $x(l) \in int(\mathbb{K})$, by (iii) of Propostion 2.7, Proposition 2.3 and Proposition 2.10, we have

$$x_{NN} - P_{x(l)_{BN}}(x(l)^{-1}_{BB}) \in int(\mathbb{K}_0),$$

where $x(l)^{-1}_{BB}$ is the inverse of $x(l)_{BB}$ in $\mathbb{V}_{BB}$. This means that

$$\begin{aligned} tr(x(l)_{NN}) &\geq tr(P_{x(l)_{BN}}(x(l)^{-1}_{BB})) = tr(x(l)^2_{BN} \circ x(l)^{-1}_{BB}) \\ &\overset{Lemma\ 6.2}{\geq} \lambda^{-1}_{max}(x(l)_{BB}, \mathbb{V}_{BB})\|x(l)_{BN}\|^2_F, \end{aligned}$$

whence

$$\|x(l)_{BN}\|^2_F \leq \lambda_{max}(x(l)_{BB}, \mathbb{V}_{BB}) tr(x(l)_{NN}) = O(\sigma(l)).$$

□

Here is one more ingredient of providing a necessary condition for the Lipschitzian error bound (LB) to hold.

Lemma 6.5. *Assume* **(C1)**, **(C2)**, **(C3)**, *and the sequence* $\{(x(l), z(l))\}_{l=1}^\infty \subseteq \mathcal{F}^\circ$. *We define*

$$\begin{aligned} \tilde{x}(l)_{BB} &= x(l)_{BB}, & \tilde{z}(l)_{BB} &= \sigma(l)^{-1} z(l)_{BB}, \\ \tilde{x}(l)_{BN} &= \sigma(l)^{-1} x(l)_{BN}, & \tilde{z}(l)_{BN} &= \sigma(l)^{-1} z(l)_{BN}, \\ \tilde{x}(l)_{NN} &= \sigma(l)^{-1} x(l)_{NN}, & \tilde{z}(l)_{NN} &= z(l)_{NN}. \end{aligned} \tag{6.12}$$

If the Lipschitzian error bound (LB) holds, then the sequence $\{(\tilde{x}(l), \tilde{z}(l))\}_{l=1}^\infty$ *is bounded.*

Proof. In light of (6.8) and (6.9), both $\{(\tilde{x}(l)_{BB}, \tilde{z}(l)_{BB})\}_{l=1}^\infty$ and $\{(\tilde{x}(l)_{NN}, \tilde{z}(l)_{NN})\}_{l=1}^\infty$ are bounded. Since the solution set $\mathcal{F}^\star$ is convex and closed by (1) of Lemma 6.1, together with (2) of Lemma 6.1 and the Lipschitzian error bound (LB), there exists a sequence $\{(x^\star(l), z^\star(l))\}_{l=1}^\infty \subseteq \mathcal{F}^\star$ such that

$$\sqrt{d_x(l) + d_z(l)} = O(\sigma(l)),$$

where

$$d_x(l) = ||x(l)_{BB} - x^\star(l)_{BB}||_F^2 + ||x(l)_{BN}||_F^2 + ||x(l)_{NN}||_F^2,$$
$$d_z(l) = ||z(l)_{BB}||_F^2 + ||z(l)_{BN}||_F^2 + ||z(l)_{NN} - z^\star(l)_{NN}||_F^2.$$

This implies that

$$||x(l)_{BN}||_F = O(\sigma(l)),\ ||z(l)_{BN}||_F = O(\sigma(l)),$$

whence $\{(\tilde{x}(l)_{BN}, \tilde{z}(l)_{BN})\}_{l=1}^\infty$ is bounded. □

We are now in a position to give a necessary condition for the Lipschitzian error bound (LB) of (SCLCP) to hold.

Theorem 6.1. *Assume* **(C1)**, **(C2)**, **(C3)**, **(C4)**, *Assumption* **A6**, *and that the sequence of strictly feasible solutions* $\{(x(l), z(l))\}_{l=1}^\infty$ *has* $\lim_{l\to\infty} \sigma(l) = 0$. *Then the Lipschitzian error bound (LB) holds only if*

$$\text{(NC1)} \qquad \kappa(P_{x(l)} P_{z(l)}) = O(1),$$

where $\kappa(P_{x(l)} P_{z(l)})$ *denotes the condition number* $||P_{x(l)} P_{z(l)}||_F ||(P_{x(l)} P_{z(l)})^{-1}||_F$ *of the product* $P_{x(l)} P_{z(l)}$.

Proof. We construct a proof by contradiction. Suppose that the Lipschitzian error bound (LB) holds but (NC1) fails. Passing to a subsequence if necessary, we may assume that

$$\kappa(P_{x(l)} P_{z(l)}) \to \infty \text{ as } l \to \infty.$$

Let $\tilde{x}(l)$ and $\tilde{z}(l)$ be specified by (6.12), then $\{(\tilde{x}(l), \tilde{z}(l))\}_{l=1}^\infty$ is bounded by Lemma 6.5. Passing to a subsequence if necessary, we may assume that $\tilde{x}(l)$ and $\tilde{z}(l)$ converges to, say, $(\tilde{x}(\infty), \tilde{z}(\infty))$.

For any given idempotent c and $x, z \in \mathbb{V}$, let

$$x = x_1 + x_{\frac{1}{2}} + x_0 \quad \text{and} \quad z = z_1 + z_{\frac{1}{2}} + z_0$$

be the Peirce decompositions of x and z with respect to c, respectively. Using (iv) and (v) of Theorem 2.1, we have

$$\begin{aligned}
&P_{x_1}\mathbb{V} \subseteq \mathbb{V}(c,1),\ \ P_{x_0}\mathbb{V} \subseteq \mathbb{V}(c,0),\\
&P_{x_1,x_{\frac{1}{2}}}\mathbb{V} \subseteq \mathbb{V}(c,1) \oplus \mathbb{V}(c,\frac{1}{2}),\\
&P_{x_0,x_{\frac{1}{2}}}\mathbb{V} \subseteq \mathbb{V}(c,0) \oplus \mathbb{V}(c,\frac{1}{2}).
\end{aligned}$$

In addition, for any $y \in \mathbb{V}$, taking into account its Peirce decomposition

$$y = y_1 + y_{\frac{1}{2}} + y_0$$

with respect to c, we have

$$P_{x_1,x_0}(y_1) = \mathbf{0} = P_{x_1,x_0}(y_0),$$

since x_1 and x_0 operator commute and $x_1 \circ x_0 = \mathbf{0}$. Then

$$P_{x_1,x_0}(y) = P_{x_1,x_0}(y_{\frac{1}{2}}) \in \mathbb{V}(c,\frac{1}{2})$$

by (ii) of Theorem 2.1. Thus

$$P_{x_1,x_0}\mathbb{V} \subseteq \mathbb{V}(c,\frac{1}{2}).$$

Consequently, we obtain

$$\begin{aligned}
&P_{x_1}P_{z_0} = \mathbf{0},\ \ P_{x_1}P_{z_1,z_0} = \mathbf{0},\ \ P_{x_1}P_{z_{\frac{1}{2}},z_0} = \mathbf{0},\\
&P_{x_0}P_{z_1} = \mathbf{0},\ \ P_{x_0}P_{z_1,z_0} = \mathbf{0},\ \ P_{x_0}P_{z_{\frac{1}{2}},z_1} = \mathbf{0},\\
&P_{x_1,x_{\frac{1}{2}}}P_{z_0} = \mathbf{0},\ \ P_{x_1,x_0}P_{z_1} = \mathbf{0},\\
&P_{x_1,x_0}P_{z_0} = \mathbf{0},\ \ P_{x_{\frac{1}{2}},x_0}P_{z_1} = \mathbf{0}.
\end{aligned} \tag{6.13}$$

Using (6.13), direct computation shows that

$$W(l) = P_{\tilde{x}(l)_{BB}}P_{\tilde{z}(l)_{BB}} + P_{\tilde{x}(l)_{BB}}P_{\tilde{z}(l)_{BN}} + 2P_{\tilde{x}(l)_{BB}}P_{\tilde{z}(l)_{BB},\tilde{z}(l)_{BN}}$$

$$
\begin{aligned}
&\sigma(l)^2(P_{\tilde{x}(l)_{BN}}P_{\tilde{z}(l)_{BB}} + P_{\tilde{x}(l)_{BN}}P_{\tilde{z}(l)_{BN}}) + P_{\tilde{x}(l)_{BN}}P_{\tilde{z}(l)_{NN}} \\
&+ 2\sigma(l)P_{\tilde{x}(l)_{BN}}(\sigma(l)P_{\tilde{z}(l)_{BB},\tilde{z}(l)_{BN}} + P_{\tilde{z}(l)_{BB},\tilde{z}(l)_{NN}} + P_{\tilde{z}(l)_{BN},\tilde{z}(l)_{NN}}) \\
&+ \sigma(l)^2P_{\tilde{x}(l)_{NN}}P_{\tilde{z}(l)_{BN}} + P_{\tilde{x}(l)_{NN}}P_{\tilde{z}(l)_{NN}} + 2\sigma(l)(P_{\tilde{x}(l)_{NN}}P_{\tilde{z}(l)_{BN},\tilde{z}(l)_{NN}} \\
&+ P_{\tilde{x}(l)_{BB},\tilde{x}(l)_{BN}}P_{\tilde{z}(l)_{BB}}) + 2\sigma(l)(P_{\tilde{x}(l)_{BB},\tilde{x}(l)_{BN}} + P_{\tilde{x}(l)_{BB},\tilde{x}(l)_{NN}} \\
&+ \sigma(l)P_{\tilde{x}(l)_{BN},\tilde{x}(l)_{NN}})P_{\tilde{z}(l)_{BN}} + 2P_{\tilde{x}(l)_{BN},\tilde{x}(l)_{NN}}P_{\tilde{z}(l)_{NN}} + 4(P_{\tilde{x}(l)_{BB},\tilde{x}(l)_{BN}} + \\
&P_{\tilde{x}(l)_{BB},\tilde{x}(l)_{NN}} + \sigma(l)P_{\tilde{x}(l)_{BN},\tilde{x}(l)_{NN}})(\sigma(l)P_{\tilde{z}(l)_{BB},\tilde{z}(l)_{BN}} + P_{\tilde{z}(l)_{BB},\tilde{z}(l)_{NN}} + P_{\tilde{z}(l)_{BN},\tilde{z}(l)_{NN}}).
\end{aligned}
$$

Since $\sigma(l) \to 0$ as $l \to \infty$, we have

$$
\begin{aligned}
W(l) \to &P_{\tilde{x}(\infty)_{BB}}P_{\tilde{z}(\infty)_{BB}} + P_{\tilde{x}(\infty)_{BB}}P_{\tilde{z}(\infty)_{BN}} + 2P_{\tilde{x}(\infty)_{BB}}P_{\tilde{z}(\infty)_{BB},\tilde{z}(\infty)_{BN}} \\
&+ P_{\tilde{x}(\infty)_{BN}}P_{\tilde{z}(\infty)_{NN}} + P_{\tilde{x}(\infty)_{NN}}P_{\tilde{z}(\infty)_{NN}} + 2P_{\tilde{x}(\infty)_{BN},\tilde{x}(\infty)_{NN}}P_{\tilde{z}(\infty)_{NN}} \\
&+ 4(P_{\tilde{x}(\infty)_{BB},\tilde{x}(\infty)_{BN}} + P_{\tilde{x}(\infty)_{BB},\tilde{x}(\infty)_{NN}})(P_{\tilde{z}(\infty)_{BB},\tilde{z}(\infty)_{NN}} + P_{\tilde{z}(\infty)_{BN},\tilde{z}(\infty)_{NN}}) \\
=:&W(\infty).
\end{aligned}
$$

Noting that

$$
\begin{aligned}
&\lambda_{min}(\sigma(l)^{-2}P_{x(l)}P_{z(l)}) = \lambda_{min}(\sigma(l)^{-2}P_{x(l)^{1/2}}P_{z(l)}P_{x(l)^{1/2}}) \\
=&\lambda_{min}(\sigma(l)^{-2}P_{P_{x(l)^{1/2}}(z(l))}) = \lambda_{min}^2(\sigma(l)^{-1}P_{x(l)^{1/2}}(z(l))),
\end{aligned}
$$

where the last equality follows from [90, Lemma 12], we obtain

$$
\lambda_{min}(W(\infty)) \geq \rho^2 > 0
$$

by Assumption (**C4**). This means that $W(\infty)$ is invertible. Consequently, $W(l)$, whence

$$
P_{x(l)}P_{z(l)} = \sigma(l)^2W(l),
$$

has bounded condition number. □

It is known (see [19]) that $\kappa(X(l)Z(l)) = O(1)$ is a necessary condition of the Lipschitzian error bound (LB) if we restrict ourselves to the primal-dual pair of SDPs. We can prove a similar result on the monotone (SCLCP) under the assumption that $\mathbb{V}$ is special.

Henceforth, we assume that $\mathbb{V}$ is special, which is derived from an associative algebra $\mathbb{A}$ with involution $*$. Moreover, for each $x \in \mathbb{V}$, by abuse of notation we still write

$$x = x_{BB} + x_{BN} + x_{BN}^* + x_{NN}$$

as its associative Peirce decomposition. Since

$$\|x_{BN}\|_F + \|x_{BN}^*\|_F \le \frac{1}{\sqrt{2}}\sqrt{\|x_{BN}\|_F^2 + \|x_{BN}^*\|_F^2},$$

together with Remark 2.2, it is readily seen that Lemmas 6.1, 6.3, 6.4 and 6.5 remain valid when stated in terms of the above partition.

Theorem 6.2. *Assume* **(C1)**, **(C2)**, **(C3)**, **(C4)**, *Assumption* **A6**, *and that the sequence of strictly feasible solutions* $\{(x(l), z(l))\}_{l=1}^{\infty}$ *has* $\lim_{l\to\infty} \sigma(l) = 0$. *Then the Lipschitzian error bound (LB) holds only if*

$$\text{(NC2)} \qquad \kappa(x(l)z(l)) = O(1),$$

where $\kappa(x(l)z(l))$ *denotes the condition number* $||x(l)z(l)||_F||(x(l)z(l))^{-1}||_F$ *of the product* $x(l)z(l)$.

Proof. We prove the statement following [19, Theorem 17] where the case of semidefinite programming was considered. Suppose that the Lipschitzian error bound (LB) holds but (NC2) fails. Passing to a subsequence if necessary, we may assume that

$$\kappa(x(l)z(l)) \to \infty \text{ as } l \to \infty.$$

Let $\tilde{x}(l)$ and $\tilde{z}(l)$ be specified by (6.12), then $\{(\tilde{x}(l), \tilde{z}(l))\}_{l=1}^{\infty}$ is bounded by Lemma 6.5. Passing to a subsequence if necessary, we may assume that $\{(\tilde{x}(l), \tilde{z}(l))\}_{l=1}^{\infty}$ converges to, say, $(x(\infty), z(\infty))$.

Using associative Peirce multiplication rules (2.7) , we have

$$\begin{aligned} v(l) =& \sigma(l)^{-1}x(l)z(l) \\ =& \tilde{x}(l)_{BB}\tilde{z}(l)_{BB} + \tilde{x}(l)_{BB}\tilde{z}(l)_{BN} + \sigma(l)\tilde{x}(l)_{BN}\tilde{z}(l)_{BN}^* + \tilde{x}(l)_{BN}\tilde{z}(l)_{BN} \\ & + \sigma(l)(\tilde{x}(l)_{BN}^*\tilde{z}(l)_{BB} + \tilde{x}(l)_{BN}^*\tilde{z}(l)_{BN} + \tilde{x}(l)_{NN}\tilde{z}(l)_{BN}^*) + \tilde{x}(l)_{NN}\tilde{z}(l)_{NN}. \end{aligned}$$

Letting l go to infinity, we obtain

$$v(l) \to \tilde{x}(\infty)_{BB}\tilde{z}(\infty)_{BB}+\tilde{x}(\infty)_{BB}\tilde{z}(\infty)_{BN}+\tilde{x}(\infty)_{BN}\tilde{z}(\infty)_{BN}+\tilde{x}(\infty)_{NN}\tilde{z}(\infty)_{NN} =: v(\infty).$$

Since

$$0 < \rho \leq \lambda_{min}(\sigma(l)^{-1}P_{x(l)^{1/2}}(z(l))) = \lambda_{min}(\sigma(l)^{-1}x(l)^{1/2}z(l)x(l)^{1/2}) = \lambda_{min}(\sigma(l)^{-1}x(l)z(l))$$

by Assumption (**C4**), $\lambda_{min}(v(\infty)) \geq \rho > 0$. Therefore $v(\infty)$ is invertible in $\mathbb{A}$. This means that $v(l)$, whence $x(l)z(l) = \sigma(l)v(l)$, has bounded condition number. □

The question now is whether the necessary condition (NC1) is equivalent to the one (NC2). The affirmative answer is given in the next theorem, whose proof hinges on the following basic observation.

Lemma 6.6. *For any $x \in \mathbb{K}$, it holds*

$$\frac{1+r}{2r}tr(x)^2 \leq tr(P_x) \leq ntr(x)^2,$$

where $r = rk(\mathbb{V})$ and $n = dim(\mathbb{V})$.

Proof. Let $x = \sum_{i=1}\lambda_i c_i$ be the spectral decomposition of x. By virtue of Lemma 12 in [90], we see that the eigenvalues of P_x have the form

$$\lambda_i\lambda_j, 1 \leq i \leq j \leq r.$$

We denote by d_{ij} the multiplicity of $\lambda_i\lambda_j$, then $1 \leq d_{ij} \leq n, 1 \leq i \leq j \leq r$. Noting that

$$tr(P_x) = \sum_{1\leq i\leq j\leq r} d_{ij}\lambda_i\lambda_j,$$

we obtain

$$\begin{aligned}&(1+\frac{1}{r})(\lambda_1+\cdots+\lambda_r)^2\\ \leq&\lambda_1^2+\cdots+\lambda_r^2+(\lambda_1+\cdots+\lambda_r)^2\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{i=1}^{r} d_{ii}\lambda_i^2 + \lambda_1(d_{11}\lambda_1 + d_{12}\lambda_2 + \cdots + d_{1r}\lambda_r) + \\
&\quad \lambda_2(d_{12}\lambda_1 + d_{22}\lambda_2 + \cdots + d_{2r}\lambda_r) + \cdots + \lambda_r(d_{1r}\lambda_1 + d_{2r}\lambda_2 + \cdots + d_{rr}\lambda_r) \\
&= 2tr(P_x) \leq n[\lambda_1^2 + \cdots + \lambda_r^2 + (\lambda_1 + \cdots + \lambda_r)^2] \\
&\leq 2n(\lambda_1 + \cdots + \lambda_r)^2 = 2ntr(x)^2.
\end{aligned}
$$

The proof is complete. □

Theorem 6.3. *For a sequence of strictly feasible solutions* $\{(x(l), z(l))\}_{l=1}^{\infty}$, *there holds*

(NC1) $$\kappa(P_{x(l)}P_{z(l)}) = O(1)$$

if and only if

(NC2) $$\kappa(x(l)z(l)) = O(1).$$

Proof. From the definition of condition number, we deduce that

$$
\begin{aligned}
&\kappa(P_{x(l)}P_{z(l)})^2 \\
=&||P_{x(l)}P_{z(l)}||_F^2 ||P_{z(l)^{-1}}P_{x(l)^{-1}}||_F^2 && \text{((vi) of Proposition 2.2)} \\
=&tr(P_{x(l)}P_{z(l)^2}P_{x(l)})tr(P_{z(l)^{-1}}P_{x(l)^{-2}}P_{z(l)^{-1}}) && \text{((iii) of Proposition 2.2)} \\
=&tr(P_{P_{x(l)}(z(l)^2)})tr(P_{P_{z(l)^{-1}}(x(l)^{-2})}) && \text{((ii) of Proposition 2.2)} \\
\leq&n^2 tr(P_{x(l)}(z(l)^2))^2 tr(P_{z(l)^{-1}}(x(l)^{-2}))^2 && \text{((iv) of Proposition 2.2 and Lemma 6.6)} \\
=&n^2 tr(x(l)^2 z(l)^2)^2 tr(z(l)^{-2}x(l)^{-2})^2 \\
=&n^2 ||x(l)z(l)||_F^4 ||(x(l)z(l))^{-1}||_F^4 \\
=&n^2 \kappa(x(l)z(l))^4,
\end{aligned}
$$

Similarly, we have

$$\kappa(P_{x(l)}P_{z(l)}) \geq \frac{r+1}{2r}\kappa(x(l)z(l))^2.$$

Consequently,

$$\kappa(P_{x(l)}P_{z(l)}) = O(1) \text{ if and only if } \kappa(x(l)z(l)) = O(1).$$

□

We now turn our attention away from necessary conditions to sufficient conditions. We will rely on the next result to study a sufficient condition of the Lipschitzian error bound (LB) for the monotone (SCLCP), whose proof follows that of [93, Lemma 2.3 and Remark 2] word-to-word and is omitted here.

Lemma 6.7. *Let $\mathfrak{L}$ be a linear subspace of $\mathbb{V}$ and $b \in \mathbb{V}$,*

$$\tilde{\mathfrak{L}} := \{x \in \mathbb{V} \mid x + \tau b \in \mathfrak{L} \text{ for some } \tau \in \mathbb{R}\}$$

be the smallest linear subspace contains $b + \mathfrak{L}$. Suppose that $\tilde{\mathfrak{L}} \cap \mathbb{K} \neq \emptyset$. Let $a^\star \in relint(\tilde{\mathfrak{L}} \cap \mathbb{K})$ and suppose without loss of generality that $a^\star$ is of the form (6.5). If $s(1), s(2), \cdots$ is a bounded sequence with

$$dist(s(l), b + \mathfrak{L}) \to 0 \text{ and } [-\lambda_{min}(s(l))]_+ \to 0 \text{ as } l \to \infty,$$

then

$$dist(s(l), (b + \mathfrak{L}) \cap \mathbb{K}) = O(\epsilon(l))$$

with

$$\epsilon(l) = dist(s(l), b + \mathfrak{L}) + [-\lambda_{min}(s(l))]_+ + ||s(l)_{BN}||_F + ||s(l)_{NN}||_F.$$

We will show next a sufficient condition for the Lipschitzian error bound (LB) under the assumption that $\mathbb{V}$ is special, which shall in turn yield Condition (NC2). Hence we shall deduce necessary and sufficient conditions for the Lipschitzian error bound (LB) in the context of a Euclidean Jordan algebra derived from an associative algebra with involution.

Theorem 6.4. *Assume* **(C1)**, **(C2)**, **(C3)**, **(C4)**, *Assumption* **A6**, *and that the sequence of strictly feasible solutions $\{(x(l), z(l))\}_{l=1}^{\infty}$ has $\lim_{l\to\infty} \sigma(l) = 0$. Then the Lipschitzian error bound (LB) holds if*

$$\text{(SC1)} \qquad ||x(l)z(l)||_F = O(\langle x(l), z(l)\rangle).$$

Proof. The proof is very similar to what the author has done for proving Theorem 20 in [19]. We prove this by contradiction. Suppose that there exists a subsequence

$\{(x(l_j), z(l_j))\}_{j=1}^{\infty}$ such that

$$\sigma(l_j)^{-1} dist((x(l_j), z(l_j)), \mathcal{F}^\star) \to \infty. \tag{6.14}$$

The assumption $\lim_{l\to\infty} \sigma(l) = 0$ and Lemma 6.4 yield that the sequence $\{(x(l), z(l))\}_{l=1}^{\infty}$ is bounded. Passing to a subsequence if necessary, we may assume that $\{(x(l_j), z(l_j))\}_{j=1}^{\infty}$ converges to, say, $(x(\infty), z(\infty))$. By Lemma 6.3, it holds

$$x(\infty)_{BB} \in int(\mathbb{K}^{(k)}) \ \text{ and } \ z(\infty)_{NN} \in int(\mathbb{K}^{(r-k)}),$$

whence

$$||x(l_j)_{BB}^{-1}||_F = O(1), \ ||z(l_j)_{NN}^{-1}||_F = O(1). \tag{6.15}$$

Since

$$\begin{aligned}
&\sigma(l_j)v(l_j) \\
=&x(l_j)_{BB}z(l_j)_{BB} + x(l_j)_{BB}z(l_j)_{BN} + x(l_j)_{BN}z(l_j)^*_{BN} + x(l_j)_{BN}z(l_j)_{NN} \\
&+ x(l_j)^*_{BN}z(l_j)_{BB} + x(l_j)^*_{BN}z(l_j)_{BN} + x(l_j)_{NN}z(l_j)^*_{BN} + x(l_j)_{NN}z(l_j)_{NN},
\end{aligned}$$

it holds

$$\sigma(l_j)v(l_j)_{BN} = x(l_j)_{BB}z(l_j)_{BN} + x(l_j)_{BN}z(l_j)_{NN}$$

by associative Peirce multiplication rules. Hence, recalling that $v(l_j) = \sigma(l_j)^{-1}x(l_j)z(l_j)$, and putting together with (SC1) , (6.15) and the first equation of (6.8), we see that

$$\begin{aligned}
&||x(l_j)_{BN}||_F \\
=&||\sigma(l_j)v(l_j)_{BN}z(l_j)_{NN}^{-1} - x(l_j)_{BB}z(l_j)_{BN}z(l_j)_{NN}^{-1}||_F && (6.16)\\
=&O(\sigma(l_j) + ||z(l_j)_{BN}||_F), && (6.17)
\end{aligned}$$

and

$$\begin{aligned}
&x(l_j)_{BN} \bullet z(l_j)_{BN} \\
=&tr(\sigma(l_j)v(l_j)_{BN}z(l_j)_{NN}^{-1}z(l_j)^*_{BN}) - tr(x(l_j)_{BB}z(l_j)_{BN}z(l_j)_{NN}^{-1}z(l_j)^*_{BN})
\end{aligned}$$

$$=O(\sigma(l_j)||z(l_j)_{BN}||_F) - ||x(l_j)_{BB}^{1/2} z(l_j)_{BN} z(l_j)_{NN}^{-1/2}||_F^2$$
$$=O(\sigma(l_j)||z(l_j)_{BN}||_F) - \Theta(||z(l_j)_{BN}||_F^2). \tag{6.18}$$

Let

$$\mathcal{G}_x := \left\{ x \in \mathbb{V} \;\middle|\; (x,z) \in \mathcal{G} \right\}, \; \mathcal{L}_{\mathcal{G}_x} := \left\{ x - x' \in \mathbb{V} \;\middle|\; (x,z),(x',z') \in \mathcal{G} \right\},$$

and

$$\mathcal{G}_z := \left\{ z \in \mathbb{V} \;\middle|\; (x,z) \in \mathcal{G} \right\}, \; \mathcal{L}_{\mathcal{G}_z} := \left\{ z - z' \in \mathbb{V} \;\middle|\; (x,z),(x',z') \in \mathcal{G} \right\},$$

then $\mathcal{L}_{\mathcal{G}_x}$, as well as $\mathcal{L}_{\mathcal{G}_z}$, forms a linear subspace of $\mathbb{V}$. We further specify the following two affine subspaces:

$$\mathfrak{A}_x := \left\{ x \in \mathbb{V} \;\middle|\; x - x^\star \in \mathcal{L}_{\mathcal{G}_x}, x_{BN} = \mathbf{0}, x_{NN} = \mathbf{0} \right\},$$
$$\mathfrak{A}_z := \left\{ z \in \mathbb{V} \;\middle|\; z - z^\star \in \mathcal{L}_{\mathcal{G}_z}, z_{BN} = \mathbf{0}, z_{BB} = \mathbf{0} \right\}.$$

We now study the following two systems:

$$\begin{cases} x \in \mathfrak{A}_x \\ x \in \mathbb{K} \end{cases} \quad \text{and} \quad \begin{cases} z \in \mathfrak{A}_z \\ z \in \mathbb{K}. \end{cases}$$

Assumption **A6**, Lemma 6.4 and $\{(x(l_j), z(l_j))\}_{j=1}^\infty \subseteq \mathcal{F}^\circ$ with $\lim_{j\to\infty} \sigma(l_j) = 0$ show that the hypotheses of Lemma 6.7 are satisfied. This implies that there exist sequences

$$\{\tilde{x}(l_j)\}_{j=1}^\infty \subseteq \mathfrak{A}_x \text{ with } \tilde{x}(l_j)_{BB} \in \mathbb{K}^{(k)}$$

and

$$\{\tilde{z}(l_j)\}_{j=1}^\infty \subseteq \mathfrak{A}_z \text{ with } \tilde{z}(l_j)_{NN} \in \mathbb{K}^{(r-k)}$$

such that

$$\sqrt{||x(l_j) - \tilde{x}(l_j)||_F^2 + ||z(l_j) - \tilde{z}(l_j)||_F^2}$$
$$\leq ||x(l_j) - \tilde{x}(l_j)||_F + ||z(l_j) - \tilde{z}(l_j)||_F \tag{6.19}$$

$$= O(||x(l_j)_{BN}||_F + ||x(l_j)_{NN}||_F + ||z(l_j)_{BN}||_F + ||z(l_j)_{BB}||_F)$$
$$\overset{(6.9),(6.17)}{=} O(\sigma(l_j) + ||z(l_j)_{BN}||_F). \tag{6.20}$$

Since $(\tilde{x}(l_j), \tilde{z}(l_j)) \in \mathcal{G}$,

$$(x(l_j) - \tilde{x}(l_j)) \bullet (z(l_j) - \tilde{z}(l_j)) = \langle x(l_j) - \tilde{x}(l_j), z(l_j) - \tilde{z}(l_j) \rangle \geq 0$$

by Remark 2.1. Thus we obtain

$$\begin{aligned}
&- 2x(l_j)_{BN} \bullet z(l_j)_{BN} \\
&\leq \langle x(l_j)_{BB} - \tilde{x}(l_j)_{BB}, z(l_j)_{BB} \rangle + \langle x(l_j)_{NN}, z(l_j)_{NN} - \tilde{z}(l_j)_{NN} \rangle \\
&\overset{(6.9)}{=} O(\sigma(l_j)(||x(l_j)_{BB} - \tilde{x}(l_j)_{BB}||_F + ||z(l_j)_{NN} - \tilde{z}(l_j)_{NN}||_F)) \\
&\overset{(6.20)}{=} O(\sigma(l_j)^2 + \sigma(l_j)||z(l_j)_{BN}||_F),
\end{aligned}$$

whence, in light of (6.18), it follows

$$||z(l_j)_{BN}||_F^2 = O(\sigma(l_j)^2 + \sigma(l_j)||z(l_j)_{BN}||_F),$$

which implies that

$$||z(l_j)_{BN}||_F = O(\sigma(l_j)).$$

Therefore, by $(\tilde{x}(l_j), \tilde{z}(l_j)) \in \mathcal{F}^\star$, it yields

$$dist((x(l_j), z(l_j)), \mathcal{F}^\star) \leq \sqrt{||x(l_j) - \tilde{x}(l_j)||_F^2 + ||z(l_j) - \tilde{z}(l_j)||_F^2} \overset{(6.20)}{=} O(\sigma(l_j)),$$

which contradicts (6.14). □

The next result is straightforward from Lemma 6.6. Its proof is very similar to what we have done for proving Theorem 6.3, and is therefore omitted here.

Corollary 6.1. *For a sequence of strictly feasible solutions* $\{(x(l), z(l))\}_{l=1}^\infty$, *there holds*

(SC1) $$||x(l)z(l)||_F = O(\langle x(l), z(l) \rangle)$$

if and only if

$$\text{(SC2)} \qquad ||P_{x(l)}P_{z(l)}||_F = O(\langle x(l), z(l)\rangle^2).$$

We summarize our discussion as a corollary.

Corollary 6.2. *Assume* **(C1)**, **(C2)**, **(C3)**, **(C4)**, *Assumption* **A6**, *and that the sequence of strictly feasible solutions* $\{(x(l), z(l))\}_{l=1}^{\infty}$ *has* $\lim_{l\to\infty} \sigma(l) = 0$. *Then the following statements are equivalent:*

$$\text{(LB)} \qquad dist((x(l), z(l)), \mathcal{F}^\star) = O(\langle x(l), z(l)\rangle);$$

$$\text{(NC1)} \qquad \kappa(P_{x(l)}P_{z(l)}) = O(1);$$

$$\text{(NC2)} \qquad \kappa(x(l)z(l)) = O(1);$$

$$\text{(SC1)} \qquad ||x(l)z(l)||_F = O(\langle x(l), z(l)\rangle);$$

$$\text{(SC2)} \qquad ||P_{x(l)}P_{z(l)}||_F = O(\langle x(l), z(l)\rangle^2).$$

Proof. Theorem 6.3 establishes that (NC1) is equivalent to (NC2), and Corollary 6.1 states that (SC1) is equivalent to (SC2). It is also known that (SC1) implies (LB) by Theorem 6.4, which in turn yields (NC2) using Theorem 6.2. Thus, it suffices to prove that (NC2) implies (SC1). We prove the statement following [19, Corollary 21].

Using Lemma 6.4 and associative Peirce multiplication rules, we obtain

$$||x(l)z(l)||_F^2 = ||\sigma(l)v(l)||_F^2 = O(\sigma(l)^2) + ||\sigma(l)v(l)_{BN}||_F^2.$$

Thus (SC1) holds if $||v(l)_{BN}||_F = O(1)$.

By (NC2), we get

$$||v(l)_{BN}||_F||(v(l)^{-1})_{BN}||_F \leq ||v(l)||_F||v(l)^{-1}||_F = \kappa(x(l)z(l)) = O(1). \tag{6.21}$$

In view of associative Peirce multiplication rules, it holds

$$\sigma(l)^{-1}(v(l)^{-1})_{BN} = (z(l)^{-1})_{BB}(x(l)^{-1})_{BN} + (z(l)^{-1})_{BN}(x(l)^{-1})_{NN}.$$

Since $x(l), z(l) \in int(\mathbb{K})$, by (iii) of Proposition 2.7 and Proposition 2.11, it implies that inverse formulae (2.8) and (2.9) are valid for $x(l)$ and $z(l)$. We denote the Schur complement of $x(l)_{BB}$ in $x(l)$ by

$$f(l) = x(l)_{NN} - x(l)^*_{BN}x(l)^{-1}_{BB}x(l)_{BN},$$

and the Schur complement of $z(l)_{NN}$ in $z(l)$ by

$$g(l) = z(l)_{BB} - z(l)_{BN}z(l)^{-1}_{NN}z(l)^*_{BN}.$$

Therefore we deduce from inverse formulae (2.8) and (2.9) that

$$\begin{aligned}
&\sigma(l)^{-1}(v(l)^{-1})_{BN} \\
=& - g(l)^{-1}x(l)^{-1}_{BB}x(l)_{BN}f(l)^{-1} - g(l)^{-1}z(l)_{BN}z(l)^{-1}_{NN}f(l)^{-1} \qquad (6.22)\\
=& - g(l)^{-1}x(l)^{-1}_{BB}(x(l)_{BN}z(l)_{NN} + x(l)_{BB}z(l)_{BN})z(l)^{-1}_{NN}f(l)^{-1} \\
=& - \sigma(l)g(l)^{-1}x(l)^{-1}_{BB}v(l)_{BN}z(l)^{-1}_{NN}f(l)^{-1}. \qquad (6.23)
\end{aligned}$$

By virtue of (iii) of Proposition 2.7 and Proposition 2.3, the Schur complements $f(l), g(l) \in \mathbb{K}$, then it follows

$$\begin{aligned}
||f(l)||_F^2 &= f(l) \bullet (x(l)_{NN} - x(l)^*_{BN}x(l)^{-1}_{BB}x(l)_{BN}) \\
&= \langle f(l), x(l)_{NN}\rangle - ||x(l)^{-1/2}_{BB}x(l)_{BN}f(l)^{1/2}||_F^2 \\
&\leq \langle f(l), x(l)_{NN}\rangle \\
&= (x(l)_{NN} - x(l)^*_{BN}x(l)^{-1}_{BB}x(l)_{BN}) \bullet x(l)_{NN} \\
&\leq ||x(l)_{NN}||_F^2.
\end{aligned}$$

Similarly, we get $||g(l)||_F \leq ||z(l)_{BB}||_F$. Thus

$$||f(l)||_F = O(\sigma(l)) \text{ and } ||g(l)||_F = O(\sigma(l))$$

by (6.9). Using (6.23), (6.8) and Lemma 2.1, we then have

$$\begin{aligned} ||v(l)_{BN}||_F &\leq \sigma(l)^{-2}||x(l)_{BB}||_F||g(l)||_F||(v(l)^{-1})_{BN}||_F||f(l)||_F||z(l)_{NN}||_F \\ &= O(||(v(l)^{-1})_{BN}||_F). \end{aligned}$$

Consequently, (6.21) yields

$$||v(l)_{BN}||_F = O(||(v(l)^{-1})_{BN}||_F^{1/2}||v(l)_{BN}||_F^{1/2}) = O(1).$$

□

We end this section by providing error bound results for a pair of primal-dual convex quadratic symmetric cone programs (P) and (D), which is straightforward from Theorem 6.1, the above corollary and the relationship between such pair of primal-dual problems and the monotone (SCLCP).

Recall that $\mathcal{O}_P$ and $\mathcal{O}_D$ are, respectively, the sets of optimal solutions for (P) and (D).

Corollary 6.3. *Assume that* **A1** *and* **A2***, the primal-dual convex quadratic symmetric cone programs (*P*) and (*D*) have strictly complementary solutions, and the sequence of primal-dual strictly feasible solutions* $\{(x(l), y(l), z(l))\}_{l=1}^{\infty}$ *of (*P*) and (*D*) has duality gaps* $\langle x(l), z(l)\rangle$ *converging to zero, and satisfies (***C4***). Then the following Lipschitzian error bound holds:*

$$\max\{dist(x(l), \mathcal{O}_P), dist(z(l), \mathcal{O}_D)\} = O(\langle x(l), z(l)\rangle)$$

only if

$$\kappa(P_{x(l)}P_{z(l)}) = O(1).$$

If, in addition, $\mathbb{V}$ *is special, then the latter is also sufficient for the Lipschitzian error* *bound and equivalent to each of the following statements:*

$$\kappa(x(l)z(l)) = O(1);$$

$$||x(l)z(l)||_F = O(\langle x(l), z(l)\rangle);$$

$$||P_{x(l)}P_{z(l)}||_F = O(\langle x(l), z(l)\rangle^2).$$

Chapter 7

Numerical results for convex quadratic semidefinite programming

For the sake of validating our approach, in this chapter we report on computational experiments of our proposed method. We present the implementation of Algorithm 4.4 in the MATLAB programming environment, and evaluate its performance regarding 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. The majority results of this chapter has been published in a peer-reviewed journal [61].

In the numerical experiments, for the sake of being comparable with those solved by the method in [101], methods used to generate data of our test problems are the same as those in [100, 101]. The following test problem E3 is used to compare the CPU time by our proposed algorithm with that by the algorithm presented in [97]. These test problems arise from the nearest correlation matrix problem, which has some instances of practical relevance, such as a 10,000 gene micro-array data set obtained from 256 drugs treated rat livers, see [74] for details. The data used for the following test problem E3 is from this dataset.

E1 $\mathcal{Q}(X) = HXH$ with $H \succ \mathbf{0}$ 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 $\mathcal{Q}(X) = H \cdot X$. The matrix H is generated randomly as follows:

`tmp=rand(n), H=0.5*(tmp+tmp').`

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.`

E3 The objective function is $\frac{1}{2}\|X - C\|_F^2$. The matrix C is obtained from the gene data sets with dimension $n = 200, 400, 800, 1000$, respectively, and has been used by the authors of [28].

We use the following three algorithms to solve each test problem:

S1 An iterative solver -Algorithm IP-QSDP [100] that uses PSQMR with constrained preconditioning. We use the code of [102].

S2 Algorithm 4.4

(A MATLAB code is available at http://clsyzx.fzu.edu.cn/linhuiling/linhuiling.html).

S3 Algorithm 2.3 presented in [97]. We use MATLAB code of [97]. We would like to mention that MATLAB code by Sun and Zhang [97] can only be used to solve Problem (1.1) for the special case where $\mathcal{L} = \tau\mathcal{I}$ with $\tau > 0$. Hence it is not available for solving test problems `E1` and `E2`.

In our implementation, we choose the following metric which possesses the desirable property that Assumption **A3** is satisfied: $(\mathcal{M}_{X,k}, \mathbf{M}_{y,k}) = (\mathcal{I} - \frac{1}{\nu_k}\mathcal{Q}, I)$ where $\nu_k > \|\mathcal{Q}\|_2$ for $k = 0, 1, \cdots$. Such choice results in $\mathcal{T}_k = \nu_k\mathcal{I}$ and $\varrho_k = 1$. The purpose of such choice is to make it easy to compute the inverse of $\mathcal{T}_k$, which in turn will make it much more efficient for solving subproblems. Of course, our choice of the metric $(\mathcal{M}_{X,k}, \mathbf{M}_{y,k})$ is just an example and there is still room for improvement in the performance of the algorithm.

Note that for any $X \in \mathbb{S}^n$, $\langle X, \mathcal{M}_{X,k}(X)\rangle - \langle X, X\rangle = -\langle \mathcal{Q}(X), X\rangle/\nu_k \leq 0$. Then $\|X\|_{\mathcal{M}_{X,k}} \leq \|X\|_F$ for any $X \in \mathbb{S}^n$ and $k \geq 0$, whence $\overline{B}(\widehat{X}^k, \hat{y}^k, \sqrt{l_d^k}) \subseteq$

$\mathscr{B}_{\mathbf{M}}(\widehat{X}^k, \hat{y}^k, l_d^k)$. It is precisely this inclusion relationship that leads us to employ the same scheme as being used in [49]. We redefine the variation

$$(var)_k = f(\widehat{X}^k, \hat{y}^k) - \min\left\{f(X,y) \;\middle|\; (X,y) \in \overline{B}(\widehat{X}^k, \hat{y}^k, 1)\right\}.$$

When f is defined on $\mathbb{R}^m$, it reduces to the variation presented in [49].

Using Corollary 4.1 and (4.13), we obtain

$$\begin{aligned}
&f(\widehat{X}^k, \hat{y}^k) - (var)_k = \min\left\{f(X,y) \;\middle|\; (X,y) \in \overline{B}(\widehat{X}^k, \hat{y}^k, 1)\right\} \\
\geq &\min\left\{f(\widehat{X}^k, \hat{y}^k) + \langle g_X^k, X - \widehat{X}^k\rangle + \langle g_y^k, y - \hat{y}^k\rangle - \tau^k \;\middle|\; (X,y) \in \overline{B}(\widehat{X}^k, \hat{y}^k, 1)\right\} \\
\geq &f(\widehat{X}^k, \hat{y}^k) - (\nu_k(\|\mathcal{M}_{X,k}(d_X^k)\|_F + \|d_y^k\|_2) + \tau^k),
\end{aligned}$$

whence $0 \leq (var)_k \leq \nu_k(\|\mathcal{M}_{X,k}(d_X^k)\|_F + \|d_y^k\|_2) + \tau^k$.

On the other hand, recalling that

$$f(\widehat{X}^k, \hat{y}^k) - \min\left\{f_{\widehat{\mathcal{W}}^k}(X,y) \;\middle|\; (X,y) \in \mathscr{B}_{\mathbf{M}}(\widehat{X}^k, \hat{y}^k, l_d^k)\right\} \leq \vartheta^k + \alpha\hat{\epsilon}^k,$$

we have

$$f(\widehat{X}^k, \hat{y}^k) - \min\left\{f(X,y) \;\middle|\; (X,y) \in \mathscr{B}_{\mathbf{M}}(\widehat{X}^k, \hat{y}^k, l_d^k)\right\} \leq \vartheta^k + \alpha\hat{\epsilon}^k.$$

Hence, for $l_d^k > 1$, $\overline{B}(\widehat{X}^k, \hat{y}^k, 1) \subseteq \mathscr{B}_{\mathbf{M}}(\widehat{X}^k, \hat{y}^k, l_d^k)$ implies that $(var)_k \leq \vartheta^k + \alpha\hat{\epsilon}^k$.

Consequently, based on the idea of [49], we replace (4.43) with

$$e_v^k > \begin{cases} \max\{\nu_k(\|\mathcal{M}_{X,k}(d_X^k)\|_F + \|d_y^k\|_2) + \tau^k, 10(\vartheta^k + \alpha\hat{\epsilon}^k)\} & \text{if } l_d^k \leq 1, \\ \min\{\nu_k(\|\mathcal{M}_{X,k}(d_X^k)\|_F + \|d_y^k\|_2) + \tau^k, \vartheta^k + \alpha\hat{\epsilon}^k\} & \text{if } l_d^k > 1, \end{cases}$$

because

$$\mathscr{B}_{\mathbf{M}}(\widehat{X}^k, \hat{y}^k, 1) \subseteq \mathscr{B}_{\mathbf{M}}(\widehat{X}^k, \hat{y}^k, l_d^k) l_d^k > 1,$$

and

$$\mathscr{B}_{\mathbf{M}}(\widehat{X}^k, \hat{y}^k, l_d^k) \subseteq \mathscr{B}_{\mathbf{M}}(\widehat{X}^k, \hat{y}^k, 1) \text{ for } l_d^k \leq 1.$$

Accordingly, we replace (4.44), (4.45) and (4.46) with

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

$$\theta_v^{k+1} = \begin{cases} \min\{\theta_v^k,\ \nu_k(\|\mathcal{M}_{X,k}(d_X^k)\|_F + \|d_y^k\|_2) + \tau^k\} & \text{if } l_d^k \le 1, \\ \min\{\theta_v^k, \vartheta^k + \alpha\hat{\epsilon}^k, \nu_k(\|\mathcal{M}_{X,k}(d_X^k)\|_F + \|d_y^k\|_2) + \tau^k\} & \text{if } l_d^k > 1, \end{cases}$$

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

respectively. Note that the above replacements do not affect the convergence of our proposed algorithm and they may force ν_k to be updated more often.

To save time, the stopping criterion in Algorithm 4.4 is replaced by

$$\tilde{\varpi}^k = \max\{l_d^k, \min\{\eta^k, \tau^k\}\} \le \varepsilon. \tag{7.1}$$

Thus (5.18) and (5.19) are reduced to

$$d((\widehat{X}^k, \hat{y}^k), \mathcal{O}_\star) \le \frac{\varepsilon}{\frac{\sigma}{\sqrt{\|D\|_F^2 + \|d\|_2^2}} - \nu_{max}\sqrt{\varepsilon}}$$

and

$$d((X^{k+1}, y^{k+1}), \mathcal{O}_\star) \le \frac{\varepsilon}{\frac{\sigma}{\sqrt{\|D\|_F^2 + \|d\|_2^2}} - \nu_{max}\sqrt{\varepsilon}},$$

respectively, which indicate that (7.1) is reasonable. We have found that, when it comes to execution time, our algorithm performs much worse without the above specializations.

The most time-consuming process of Algorithm 4.4 is the eigenvalue estimation. Considering that eigenvalue problems in question are not necessarily sparse, we adopt the existing program **eigifp** [45] to solve them, whose underlying algorithm is based on an inverse free preconditioned Krylov subspace projection method [30].

In step 2 of Algorithm 4.4, we need to compute the subproblem (SQSCP), where the updating rule for $\widehat{\mathcal{W}}^k$ is identical to the one presented in [38]. With the help of the symmetrized Kronecker product [3], using the analogous technique as that presented in [38], we can cast (SQSCP) as a quadratic semidefinite program with small size. We solve the quadratic semidefinite program by a feasible primal-dual interior-point method, whose essential structure is that of Mehrotra-type predictor-corrector primal-dual path-following algorithm in [101]. The method used here differs in that we should guarantee the feasibility at each iteration. The Nesterov-Todd symmetrization scheme [99] is employed.

The initial iterate for Algorithm 4.4 is given by $X^0 = \mathbf{0}$, $y^0 = \mathbf{0}$. The value of parameters present in the algorithm are listed as follows:

$$\begin{aligned}
&(\text{the upper bound of the number of columns of all } P_k s) \leq 30,\\
&m_L = 0.1, m_R = 0.55, \delta^0 = 10^{-6}, \theta_v^0 = 10^{15},\\
&\nu_{min} = \begin{cases} \|\mathcal{Q}\|_2 + 0.01 & \text{if } \|\mathcal{Q}\|_2 > 1,\\ 1.01\|\mathcal{Q}\|_2 & \text{otherwise,}\end{cases}\\
&\nu_0 = \max\left\{\nu_{min}, \sqrt{\|\mathcal{A}(\overline{W}_0) - b\|_2^2 + \|\mathcal{Q}(\overline{W}_0)\|_F^2}\right\},
\end{aligned}$$

where $\overline{W}_0 = \alpha \hat{p}_{ap}^0 (\hat{p}_{ap}^0)^T$, and $\hat{p}_{ap}^0$ is a normalized approximate eigenvector corresponding to the approximate maximum eigenvalue of $-C - \mathcal{A}^*(y^0) - \mathcal{Q}(X^0)$. The stopping criterion is based on (7.1) with

$$\varepsilon = \begin{cases} 10^{-4} & \text{for test problems in } \texttt{E1} \text{ and } \texttt{E3},\\ 10^{-3} & \text{for test problems in } \texttt{E2},\end{cases}$$

in the implementation. We will see that, in the report on the performance of Algorithm S1 and S2, $\mathcal{Q}(X) = HXH$ for some $H \succ \mathbf{0}$ is a special case of $\mathcal{Q}(X) = K \cdot X$ for some rank-one matrix $K \succeq \mathbf{0}$. To save time and memory, we set `E2` to a lower accuracy.

The performance results are given in Table 7.1 and Table 7.2, where we compare the proposed inexact spectral bundle method `S2` to algorithms `S1` and `S3`. The columns corresponding to "iter" give the total number of iterations. The column corresponding to "value(p)" gives the minimum value of Problem (1.1) computed by Algorithm `S1`. Under Assumptions **A1**, Problem (3.5) is the negative dual of Problem (1.1) without the constant $\langle B, \mathcal{L}^2(B)\rangle/2$, so we refer to the column "value(d)" as the negative minimum value of Problem (3.5) solved by Algorithm `S2`, which makes it easy to compare. Note that both two quantities discard the constant $\langle B, \mathcal{L}^2(B)\rangle/2$. The columns corresponding to "ϕ", "ν" and "ϖ" present the accuracy measure defined by (46) in [100], (7.1) and [97], 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 the elapsed time in seconds. Those entries being "err" mean that the algorithm terminates with out-of-memory errors.

Table 7.1: Performance of Algorithms 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	14	58	-2.42e+5	1.33e-7	71	26	104.2	-2.45e+5	0.00668	0.0317	8.48e-5
400	13	246.2	-5.19e+5	1.76e-7	67	27	187.3	-5.30e+5	0.0124	0.0464	9.28e-5
800	15	1777.8	-1.05e+6	1.82e-7	85	26	895.8	-1.07e+6	0.0226	0.127	8.35e-5
1600	17	18714.2	-2.64e+6	3.04e-7	132	44	7350	-2.74e+6	0.0421	0.0218	8.99e-5
2000	err	err	err	err	108	39	11042.7	-2.99e+6	0.0459	0.0775	6.49e-5
E2 200	12	16.3	-2.3e+2	1.84e-7	473	78	812.3	-3.3e+2	0.414	0.654	7.66e-4
400	12	96.3	-4.58e+2	5.57e-7	487	119	2032.8	-9.93e+2	0.571	0.752	7.64e-4
800	12	579.6	-9.24e+2	8.59e-7	469	76	5423.5	-2.73e+3	0.874	0.499	9.63e-4

Table 7.2: Performance of Algorithms S1, S2 and S3 on the problem set E3

	S1			S2				S3		
n	iter	time(s)	ϕ	iter	SS	time(s)	ν	iter	time(s)	ϖ
E3 200	9	21	9.26e-7	165	6	215.8	9.13e-5	101	34.5	5.043e-2
400	10	100.9	5.64e-7	219	2	345.8	2.68e-6	101	213.2	1.869e-2
800	10	622.8	7.20e-7	134	2	709.1	1.97e-5	101	1775	3.233e-2
1000	11	4944.5	3.21e-7	457	2	3740.9	5.47e-5	101	3207.7	9.206e-3

Table 7.1 and Table 7.2 provide some information that we sum up below.

1. As a whole, Algorithm S2 outperforms Algorithm S1 on the problem set E1. For problems with size of no less than 400, our algorithm consumes less CPU time than Algorithm S1. In particular, we see that our method can find reasonably accurate solution in each of the three problems E1-400, E1-800 and E1-1600 in about half the time required by Algorithm S1. We explain the case when $n = 2000$ in item 3.

2. For problems with size of more than 400, the CPU time by our proposed algorithm is in between the CPU time by Algorithm S1 and that by Algorithm S3 in [97], which is evident in Table 7.2. At the same time, we see that Algorithm S3 terminates in E3 since it reaches the maximum number of iterations allowed.

3. Algorithm S1 is not able to solve the test problems in E1-2000 since interior-point methods are not applicable because of memory requirements, while Algorithm S2 performs well on these problems. This demonstrates that bundle methods may 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. It should be clear from the examples of Table 7.1 that Algorithm S2 uses much more time than Algorithm S1. Furthermore, the rather large norm of the subgradient shows that the values cannot be expected to be "good" approximations of the objective function, which also is illustrated by comparing two columns "value(d)" and "value(p)" of the problem set E2. We will improve the performance of our algorithm in the case when $\mathcal{Q}(X) = H \cdot X$ in the future.

5. From the column "time(s)" of Algorithm S2 in Table 7.1, we see that the computation time of E1 is much smaller than that of E2. This arises from the fact that $\mathcal{Q}(X) = HXH$ with $H \succ \mathbf{0}$ can be viewed as a special case of $\mathcal{Q}(X) = H \cdot X$. Indeed, using the same technique as used in linear semidefinite programs [35, 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:

$$\widetilde{X} = Q^T XQ,\ \widetilde{Z} = Q^T ZQ,\ \widetilde{W} = Q^T WQ, \widetilde{C} = Q^T CQ,$$
$$\widetilde{\mathcal{A}}(\tilde{\cdot}) = (\langle Q^T A_1 Q, \tilde{\cdot}\rangle, \cdots, \langle Q^T A_m Q, \tilde{\cdot}\rangle)^T, \widetilde{\mathcal{Q}}(\tilde{\cdot}) = Q^T HQ(\tilde{\cdot})Q^T HQ,$$

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

Chapter 8

Conclusions and future work

In this thesis, we have proposed an inexact spectral bundle method for solving linearly constrained convex quadratic symmetric cone 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 method finds two sequences $\{(x^k, y^k)\}_{k=1}^{\infty}$ and $\{w^k\}_{k=1}^{\infty}$, simultaneously. When $\delta^k \downarrow \delta > 0$, limit points of both sequences are δ^{F_1}**-optimal** solutions for the eigenvalue minimization problem (EigForm) and for the CQSCP (P), respectively. In addition, a Lipschitzian error bound for the eigenvalue minimization problem (EigForm) is given under a Slater type condition, which also demonstrates the practicality of the stopping rule for Algorithm 4.4.

Likewise, we have studied the Lipschitizan error bound

$$\max\{dist(x(l), \mathcal{O}_P), dist(z(l), \mathcal{O}_D)\} = O(\langle x(l), z(l) \rangle)$$

with respect to the sequence $\{(x(l), y(l), z(l))\}_{l=1}^{\infty}$ that lies within a wide neighborhood of the primal-dual central path of a convex quadratic symmetric cone programming problem. We put ourselves in a more general setting–a framework of a monotone symmetric cone linear complementarity problem–to achieve our error bound results. Under assumptions of strict complementarity and Slater's condition, we have established that if the underlying Euclidean Jordan algebra is special, then the Lipschitzian error bound can be deduced by either of the following statements, and vice versa.

1. The condition number of $P_{x(l)}P_{z(l)}$ is bounded.

2. The condition number of $x(l)z(l)$ is bounded.

3. The Frobenius norm of the product $x(l)z(l)$ grows at most linearly with the trace of the same product.

4. The Frobenius norm of the product $P_{x(l)}P_{z(l)}$ grows at most quadratically with the trace of $x(l)z(l)$.

It has been shown that the first result furnishes a necessary condition for the Lipschitzian error bound in the context of general Euclidean Jordan algebras. For future research, it is of interest to ask whether this condition is sufficient for the Lipschitzian error bound.

When applying Algorithm 4.4 to solve a convex quadratic semidefinite programming problem, we see that it does require first-order information and is easy to implement. Numerical experiments on a set of the nearest correlation matrix problem with matrices of order up to 2000 show that 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, for example, [44].

Bibliography

[1] A. A. Albert. A structure theory for Jordan algebras. *Ann. of Math.*, 48(2):546–567, 1948.

[2] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Math. Program., Ser. B*, 95(1):3–51, 2003.

[3] 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.

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

[5] 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.

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

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

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

[9] D. Benterki and A. Keraghel. Finding a strict feasible solution of a linear semidefinite program. *Appl. Math. Comput.*, 217(13):6437–6440, 2011.

[10] P. Bonami and M. A. Lejeune. An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Oper. Res.*, 57(3):650–670, 2009.

[11] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal. *Numerical Optimization. Theoretical and Practical Aspects*. Universitext. Springer, Berlin, 2nd edition, 2006.

[12] 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.

[13] J. V. Burke and S. Deng. Weak sharp minima revisited. I. Basic theory. *Control Cybernet.*, 31(3):439–469, 2002.

[14] J. V. Burke and S. Deng. Weak sharp minima revisited. II. Application to linear regularity and error bounds. *Math. Program.*, 104(2-3):235–261, 2005.

[15] J. V. Burke and S. Deng. Weak sharp minima revisited. III. Error bounds for differentiable convex inclusions. *Math. Program.*, 116(1-2):37–56, 2009.

[16] J. V. Burke and M. C. Ferris. Weak sharp minima in mathematical programming. *SIAM J. Control Optim.*, 31(5):1340–1359, 1993.

[17] J. S. Chen and S. Pan. An entropy-like proximal algorithm and the exponential multiplier method for convex symmetric cone programming. *Comput. Optim. Appl.*, 47(3):477–499, 2010.

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

[19] C. B. Chua. Analyticity of weighted central path and error bounds for semidefinite programming. *Math. Program.*, 115:239–271, 2008.

[20] P. P. B. Eggermont. Multiplicative iterative algorithms for convex programming. *Linear Algebra Appl.*, 130:25–42, 1990.

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

[22] F. Facchinei and J. S. Pang. *Finite-dimensional variational inequalities and complementarity problems. Vol. I.* Springer-Verlag, New York, USA, 2003.

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

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

[25] L. Faybusovich. Jordan-algebraic approach to convexity theorems for quadratic mappings. *SIAM J. Optim.*, 17(2):558–576, 2006.

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

[27] 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.

[28] Y. Gao and D. Sun. A majorized penalty approach for calibrating rank constrained correlation matrix problems. *Working paper, Department of Mathematics, National University of Singapore, Singapore*, May 2010. http://www.math.nus.edu.sg/ matsundf/MajorPen.pdf.

[29] F. Giannessi and A. Maugeri. *Variational Analysis and Applications. Nonconvex Optimization and its Applications, Vol. 79.* Springer-Verlag, New York, USA, 2005.

[30] 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.

[31] M. S. Gowda and R. Sznajder. Schur complements, Schur determinantal and Haynsworth inertia formulas in Euclidean Jordan algebras. *Linear Algebra Appl.*, 432:1553–1559, 2010.

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

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

[34] C. Helmberg. SBmethod–a C++ implementation of the spectral bundle method. Manual to Version 1.1. *ZIB-Report ZR-00-35, Konrad-Zuse-Zentrum Berlin*, October 2000. http://www-user.tu-chemnitz.de/∼helmberg/SBmethod/.

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

[36] C. Helmberg and K. C. Kiwiel. A spectral bundle method with bounds. *Math. Program.*, 93(2):173–194, 2002.

[37] C. Helmberg and F. Oustry. Bundle methods to minimize the maximum eigenvalue function. In *Handbook of semidefinite programming*, volume 27 of *Internat. Ser. Oper. Res. Management Sci.*, pages 307–337. Kluwer Acad. Publ., 2000.

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

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

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

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

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

[43] U. Hirzebruch. Der min-max-satz von E. Fischer für formal-reelle Jordan-Algebren. *Math. Ann.*, 186:65–69, 1970.

[44] 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.

[45] 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.

[46] 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.

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

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

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

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

[51] M. Kojima, M. Shida, and S. Shindoh. A predictor-corrector interior-point algorithm for the semidefinite linear complementarity problem using the Alizadeh-Haeberly-Overton search direction. *SIAM J. Optim.*, 9(2):444–465, 1999.

[52] M. Kojima, M. Shida, and S. Shindoh. Search directions in the SDP and the monotone SDLCP: generalization and inexact computation. *Math. Program., Ser. A*, 85(1):51–80, 1999.

[53] 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.

[54] M. Kojima, S. Shindoh, and S. Hara. Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM J. Optim.*, 7(1):86–125, 1997.

[55] L. Kong, L. Tunçel, and N. Xiu. Equivalent conditions for Jacobian nonsingularity in linear symmetric cone programming. *J. Optim. Theory Appl.*, 148:364–389, 2011.

[56] M. Konstantin and N. Alantha. Complex semidefinite programming revisited and the assembly of circular genomes. In *ICS 2011*, pages 444–459. Tsinghua University Press, 2011. http://conference.itcs.tsinghua.edu.cn/ICS2011/content/papers/27.html.

[57] A. Ya. Kruger. Generalized differentials of nonsmooth functions, and necessary conditions for an extremum. *Sibirsk. Mat. Zh.*, 26(3):78–90, 1985.

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

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

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

[61] H. L. Lin. An inexact spectral bundle method for convex quadratic semidefnite programming. *Comput. Optim. Appl.*, in press, DOI: 10.1007/s10589-011-9443-x:1–45, 2011.

[62] A. Ling. Approximation algorithms for Max 3-section using complex semidefinite programming relaxation. In *Combinatorial optimization and applications*, volume 5573 of *Lecture Notes in Comput. Sci.*, pages 219–230. Springer, 2009.

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

[64] Z. Lu and R. D. C. Monteiro. Limiting behavior of the Alizadeh-Haeberly-Overton weighted paths in semidefinite programming. *Optim. Methods Softw.*, 22(5):849–870, 2007.

[65] I. J. Lustig. Feasibility issues in a primal-dual interior-point method for linear programming. *Math. Program., Ser. A.*, 49(2):145–162, 1990/91.

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

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

[68] K. McCrimmon. *A Taste of Jordan Algebras*. Springer-Verlag, New York, USA, 2004.

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

[70] 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.

[71] H. D. Mittelmann. An independent benchmarking of SDP and SOCP solvers. *Math. Program.*, 95(2):407–430, 2003.

[72] M. M. Moldovan. *A Geršgorin type theorem, spectral inequalities, and simultaneous stability in Euclidean Jordan algebras*. PhD thesis, University of Maryland, Baltimore County, 2009.

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

[74] G. Natsoulis, C. I. Pearson, J. Gollub, B. P. Eynon, J. Ferng, R. Nair, R. Idury, M. D. Lee, M. R. Fielden, R. J. Brennan, A. H. Roter, and K. Jarnagin. The liver pharmacological and xenobiotic gene response repertoire. *Mol. Syst. Biol.*, 4:1–12, 2008.

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

[76] H. V. Ngai, D. T. Luc, and M. Théra. Extensions of Fréchet ε-subdifferential calculus and applications. *J. Math. Anal. Appl.*, 268:266–290, 2002.

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

[78] 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.

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

[80] M. L. Overton and R. S. Womersley. Second derivatives for optimizing eigenvalues of symmetric matrices. *SIAM J. Matrix Anal. Appl.*, 16(3):697–718, 1995.

[81] J. S. Pang. Methods for quadratic programming: A survey. *Comput. Chem. Eng.*, 7(5):583–594, 1983.

[82] J. S. Pang. Error bounds in mathematical programming. *Math. Program.*, 79(1-3):299–332, 1997.

[83] 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.

[84] H.-D. Qi and D. Sun. An augmented Lagrangian dual approach for the H-weighted nearest correlation matrix problem. *IMA J. Numer. Anal.*, 31(2):491–511, 2011.

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

[86] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, 14(5):877–898, 1976.

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

[88] 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.

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

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

[91] 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.

[92] M. V. Solodov. A bundle method for a class of bilevel nonsmooth convex minimization problems. *SIAM J. Optm.*, 18(1):242–259, 2007.

[93] J. F. Sturm. Error bounds for linear matrix inequalities. *SIAM J. Optim.*, 10:1228–1248, 2000.

[94] J. F. Sturm and S. Zhang. On sensitivity of central solutions in semidefinite programming. *Math. Program.*, 90(2):205–227, 2001.

[95] D. Sun and J. Sun. Löwner's operator and spectral functions in Euclidean Jordan algebras. *Math. Oper. Res.*, 33(2):421–445, 2008.

[96] J. Sun, K. C. Toh, and G. Zhao. An analytic center cutting plane method for semidefinite feasibility problems. *Math. Oper. Res.*, 27(2):332–346, 2002.

[97] 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.

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

[99] 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.

[100] K. C. Toh. An inexact primal-dual path-following algorithm for convex quadratic SDP. *Math. Program.*, 112:221–254, 2008.

[101] 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.

[102] 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.

[103] Z. Wu and J. J. Ye. First-order and second-order conditions for error bounds. *SIAM J. Optim.*, 14:621–645, 2003.

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

[105] A. Yoshise. Interior point trajectories and a homogeneous model for nonlinear complementarity problems over symmetric cones. *SIAM J. Optim.*, 17(4):1129–1153, 2006.

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

[107] 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.