Some Properties of the Augmented Lagrangian in Nonlinear Semidefinite Optimization*

JIE SUN[1], **LI-WEI ZHANG**[2] and **YUE WU**[3]

[1]*Department of Decision Sciences and Singapore-MIT Alliance, National University of Singapore, Republic of Singapore 119260(e-mail:jsun@nus.edu.sg);*

[2]*Department of Applied Mathematics, Dalian University of Technology, Dalian 116024, China; CAIS, School of Computer Engineering, Nanyang Technological University, Republic of Singapore 639798(e-mail:aslwzhang@ntu.deu.sg);*

[3]*School of Management, University of Southampton, Highfield Southampton, SO17 1BJ, United Kingdom. (e-mail: y.wu@soton.ac.uk).*

Abstract. We study the properties of the augmented Lagrangian function for nonlinear semidefinite programming. It is shown that under a set of sufficient conditions the augmented Lagrangian algorithm is locally convergent when the penalty parameter is larger than a threshold. An error estimate of the solution, depending on the penalty parameter, is also established.

Key words: semidefinite programming, augmented Lagrangian.

1 Introduction

The augmented Lagrangian algorithm is a classical method for solving nonlinear programming problems. At the core of the algorithm is the augmented Lagrangian function defined by [17]

$$F(x,u,\tau) = f(x) + \frac{1}{2\tau}\sum_{i=1}^{m}\Big[[u_i - \tau c_i(x)]_+^2 - u_i^2\Big] \quad \tau > 0, \tag{1.1}$$

where $f : \mathbf{R}^n \to \mathbf{R}$, $c_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1,...,m$ are twice continuously differentiable; $u = (u_1,...,u_m)^T \in \mathbf{R}^m$ and $[\cdot]_+ = \max\{\cdot, 0\}$. The augmented Lagrangian (1.1) corresponds to the inequality-constrained nonlinear program

$$\min\ f(x) \text{ subject to } c_i(x) \geq 0, i = 1,...,m. \tag{1.2}$$

A different augmented Lagrangian can be defined to accommodate both inequality and equality constraints, but for simplicity we concentrate on the inequality-constrained case. It is shown by Rockafellar [17] that under a set of conditions (e.g., the quadratic growth condition, τ being large, and the second order sufficiency conditions) the augmented Lagrangian algorithm

$$x^{k+1} = \text{argmin}_x L(x,u^k,\tau) \tag{1.3}$$

$$u^{k+1} = u^k + \tau\nabla_u F(x^{k+1},u^k,\tau) \tag{1.4}$$

*The first author is partially supported by Singapore-MIT Alliance and Grants RP314000-028/042/057-112 of National University of Singapore. The second author is partially supported by the Funds of Ministry of Education of China for PhD Units under No. 20020141013 and the National Natural Science Foundation of China under project grant No.10471015.

will generate a sequence $\{(x^k, u^k)\}$ converging to a local optimal pair (x^*, u^*) at a linear rate, therefore providing an algorithm for the nonlinear program (1.2).

Consider the following semidefinite optimization problem

$$\begin{aligned} \min_{x\in\mathbf{R}^n} \quad & f(x) \\ \text{s.t.} \quad & C(x) \succeq 0, \end{aligned} \tag{1.5}$$

where $f : \mathbf{R}^n \mapsto \mathbf{R}$ and $C : \mathbf{R}^n \mapsto \mathcal{S}^m$. We denote by $\mathcal{S}^m$ the space of $m \times m$ real symmetric matrices equipped with the inner product $A \bullet B = \mathrm{Tr}(AB)$ ("Tr" stands for the trace) and assume that both functions are twice continuously differentiable. For $A \in \mathcal{S}^m$, the notation $A \succeq 0 (A \succ 0)$ means that the matrix A is positive semidefinite (positive definite). We will also use $\mathcal{S}^m_+$ and $\mathcal{S}^m_{++}$ to denote the cones of positive semidefinite and positive definite matrices, respectively.

Obviously, the nonlinear program (1.2) is a special case of (1.5), in which $C(x) = \mathrm{diag}\,(c_1(x), ..., c_m(x))$, where "diag" denotes a diagonal matrix. On the other hand, the constraint $C(x) \succeq 0$ is of course more general than the constraints of (1.2). One may think of writing $C(x) \succeq 0$ as m constraints of the form

$$\lambda_1(C(x)) \geq 0, ..., \lambda_m(C(x)) \geq 0,$$

where $\lambda_i(C(x)),\ i = 1..., m$ represents the eigenvalues of the matrix $C(x)$. However, it is well known that eigenvalues are non-differentiable functions of the matrix. Thus, such an idea does not lead to a direct application of the augmented Lagrangian theory of nonlinear programming. New analysis of the augmented Lagrangian is therefore necessary.

Analogously, we define the augmented Lagrangian of the problem (1.5) as

$$F(x, U, \tau) = f(x) + \frac{1}{2\tau}\left[\mathrm{Tr}[U - \tau C(x)]_+^2 - \mathrm{Tr}(U^2)\right] \quad \tau > 0. \tag{1.6}$$

In this context, $[\cdot]_+$ denotes the matrix projection to $\mathcal{S}^m_+$. Namely, let $A = P\,\mathrm{diag}\,(\lambda_1, \ldots, \lambda_m)P^T$ be an orthogonal decomposition of $A \in \mathcal{S}^m$. Then

$$A_+ = P\,\mathrm{diag}\ ([\lambda_1]_+, \ldots, [\lambda_m]_+)\, P^T.$$

This function is well defined and is independent of the choice of P. Similarly the matrix absolute value is defined as

$$|A| = P\,\mathrm{diag}\,(|\lambda_1|, \ldots, |\lambda_m|)P^T.$$

This paper is devoted to the discussion of the properties of $F(x, U, \tau)$ that are related to an augmented Lagrangian algorithm for (1.5). Semidefinite optimization problems (SOPs) arise naturally in a wide range of applications in engineering, optimal control, statistics, combinatorial optimization, and eigenvalue optimization, see, e.g., [1, 14, 22, 23] and the references therein. While there has been a large number of research articles addressing linear SOPs since 1990s, only few papers considered theory and algorithms for the nonlinear case in which f might be nonconvex and C might be nonlinear. Among them, Lewis and Overton [12] and Lewis [11] considered nonsmooth analysis and optimization problems involving eigenvalues. Bonnans and Shapiro [3] and Shapiro [18] established first and second order optimality and perturbation analysis for SOP. Forsgren [7] discussed optimality conditions for nonconvex semidefinite programming. Sun and Sun [20] presented a general theory of semismooth matrix functions and proved that the matrix absolute-value function is strongly semismooth. Chen, Qi and Tseng [4] obtained important results in nonsmooth analysis of matrix functions. Pang, Sun and Sun [16] developed a strong stability theory for semidefinite variational inequality problems. Shapiro and Sun [19] discussed duality properties of the augmented Lagrangian. There are also papers involving applications of nonlinear SOP and their algorithms [5, 6, 14, 15, 21, 24]. However, much work is yet to be done to effectively solve nonlinear semidefinite optimization problems including (1.5).

In this paper we discuss algorithm-related properties of the augmented Lagrangian (1.6). We provide a set of assumptions that guarantee a sequence generated by a similar algorithm to (1.3)-(1.4) to converge to a local minimizer of the problem (1.5). We also estimate the rate of convergence.

The following notation and terminology are used throughout the paper. Let $\mathbf{R}^{m\times n}$ denote the space of $m \times n$ matrices. The symbol "$\otimes$" denotes the Kronecker product of matrices

$$A \otimes B = \begin{bmatrix} a_{11}B, & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{n1}B, & \cdots & a_{nn}B \end{bmatrix}.$$

It is easy to see that $\otimes$ is a linear operator (see, e.g., [8] for more basic properties of the Kronecker product). For an $m \times n$ matrix B we use $\mathrm{vec}(B)$ to denote the $mn \times 1$ vector obtained by stacking columns of B in the natural order. It can be shown that $\mathrm{vec}\,(ABC) = (C^T \otimes B)\mathrm{vec}\,(A)$. The notation $\nabla C(x)$ represents the F(réchet)-derivative (i.e. the gradient) of the mapping $C(\cdot)$ at x. Thus, $\nabla C(x)$ is a linear operator from $\mathbf{R}^n$ into $\mathcal{S}^m$ defined by

$$[\nabla C(x)]y = \sum_{i=1}^{n} y_i C_i(x)$$

where $C_i(x) = \partial C(x)/\partial x_i$ is the $m \times m$ partial derivative matrix. The matrix form of $\nabla C(x)$ is written as $\partial C(x)/\partial x$, denoting the $m^2 \times n$ Jacobian of the mapping $\mathrm{vec}\, C(\cdot)$, i.e.

$$\partial C(x)/\partial x = [\mathrm{vec}\, C_1(x), \ldots, \mathrm{vec}\, C_n(x)].$$

Let

$$C_{ij}(x) = \partial^2 C(x)/\partial x_i \partial x_j \quad i, j = 1, ..., n$$

be the $m \times m$ second order partial derivative matrices, see [23]. A diagonal matrix is written as $\mathrm{diag}\,(\alpha_1, ..., \alpha_k)$ whereas a block diagonal matrix is written as $\mathrm{diag}\,(A_1, ..., A_k)$. We use $\mathcal{D}_r$ to represent the set of $r \times r$ real diagonal matrices with non-increasing diagonal entries, and $\mathcal{O}$ to represent the set of $m \times m$ orthogonal matrices. Finally,

$$B(U^*, \epsilon) = \{U | \|U - U^*\|_F < \epsilon\},$$

where $\|\cdot\|_F$ denotes the Frobenius norm, i.e. $\|A\|_F = (A \bullet A)^{1/2}$.

The paper is organized as follows. We state assumptions, calculate the gradient and the Hessian of $F(x, U, \tau)$ with respect to x, and prove that x^* is a strict local minimizer of $F(x, U^*, \tau)$ under our assumptions in Section 2. We present an augmented Lagrangian algorithm and prove a local convergence theorem, which shows that when τ is large enough and U^k is close to the Lagrange multiplier U^*, the sequence $\{x^k\}$ linearly converges to a local minimum point of problem (1.5) in Section 3.

2 Properties of $F(x, U, \tau)$

2.1 Basic assumptions

Some basic assumptions are made on Problem (1.5). Let $L(x, U) = f(x) - U \bullet C(x)$ be the Lagrange function of the problem.
Assumptions.

(i) Problem (1.5) has a local minimum x^*.

(ii) Robinson's constraint qualification holds at x^*. That is

$$\exists\, h \in \mathbf{R}^n : \quad C(x^*) + \nabla C(x^*)h \succ 0.$$

According to Theorem 5.84 of [3], Assumptions (i) and (ii) imply that the set of Lagrange multiplier matrices is nonempty and bounded. Therefore, there exists a matrix $U^* \in S_m$ such that (x^*, U^*) is a Karush-Kuhn-Tucker (KKT) pair of the problem (1.5). That is, (x^*, U^*) satisfies

$$\nabla_x L(x^*, U^*) = 0,$$

$$U^* C(x^*) = 0,$$

$$U^* \in S_m^+, \quad C(x^*) \in S_m^+.$$

(iii) The strict complementarity condition holds at (x^*, U^*). i.e.

$$\operatorname{rank}(C(x^*)) = r, \quad \operatorname{rank} U^* = m - r.$$

(iv) The Lagrangian Hessian $\nabla^2_{xx} L(x^*, U^*)$ is positive definite on the subspace $\mathcal{C}(x^*) = \{y \in \mathbf{R}^n | \sum_{i=1}^n y_i E^T C_i(x^*) E = 0\}$, where E is an $m \times (m - r)$ matrix whose columns form a basis for the null space of the matrix $C(x^*)$.

Remark 1 *The reader who is familiar with the so-called critical cone may find that the subspace $\mathcal{C}(x^*)$ in (iv) is in fact the critical cone of problem (1.5) since under the strict complementarity condition the critical cone is a linear subspace, e.g., see [3, p.490]. The main role of Assumption (iv) in this paper is to show the positive definiteness of $\nabla^2_{xx} F(x, U, \tau)$ in a neighborhood of (x^*, U^*) that will be used in the proof of Proposition 8 and Theorem 11 in particular. In addition, Assumptions (i), (ii), and (iv) guarantee that x^* is a strict local minimum of (1.5), see [7, Theorem 3].*

2.2 The first order differential of $F(x, U, \tau)$

The following lemmas are useful in discussing differential properties of $F(x, U, \tau)$. To simplify the discussion, we introduce some notations as follows.

$$M(x, U, \tau) = U - \tau C(x), \quad M_+(x, U, \tau) = [U - \tau C(x)]_+, \quad M_+^2(x, U, \tau) = [M_+(x, U, \tau)]^2.$$

Lemma 2 *If Assumptions (i)-(iii) hold, then there exist a matrix $Q \in \mathcal{O}$, an $r \times r$ diagonal matrix Γ_1, and an $(m - r) \times (m - r)$ diagonal matrix Γ_2 such that*

$$U^* = Q \begin{bmatrix} \Gamma_1 & 0 \\ 0 & 0 \end{bmatrix} Q^T \quad C(x^*) = Q \begin{bmatrix} 0 & 0 \\ 0 & \Gamma_2 \end{bmatrix} Q^T, \tag{2.3}$$

and $\Gamma_1 \in \mathcal{D}_r$, $\Gamma_2 \in \mathcal{D}_{m-r}$ with

$$\begin{bmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{bmatrix} \succ 0. \tag{2.4}$$

Consequently, $M(x^*, U^*, \tau)$is nonsingular for any $\tau > 0$ and there hold

$$M(x^*, U^*, \tau) = Q \begin{bmatrix} \Gamma_1 & 0 \\ 0 & -\tau\Gamma_2 \end{bmatrix} Q^T \quad \textit{and} \quad M_+(x^*, U^*, \tau) = U^*.$$

Proof. Since $U^* \in \mathcal{S}_+^m$, $C(x^*) \in \mathcal{S}_+^m$, and $U^* C(x^*) = 0$, the two matrices can be simultaneously diagonalized. This, together with Assumption (iii), implies (2.3) and (2.4). The rest of the lemma is obvious from (2.4). $\square$

The key term in $F(x, U, \tau)$ is $\mathrm{Tr}[M_+^2(x, U, \tau)]$, which is actually a *spectral function* studied in Lewis [10] and Lewis and Sendov [13]. Let $Y \in \mathcal{S}^m$ and let $\lambda_1(Y) \geq \lambda_2(Y) \geq \cdots \geq \lambda_m(Y)$ be its eigenvalues in decreasing order. Let $\lambda(Y) = (\lambda_1(Y), \lambda_2(Y), \ldots, \lambda_m(Y))^T$ and let $g : \mathbf{R}^m \to \mathbf{R}$ be a *symmetric function* in the sense that the value of the function is invariant under permutation of its variables.

Lemma 3 *([13, Lemma 2.1]) Let $Y = Q diag \lambda(Y) Q^T$ be an orthogonal decomposition of $Y \in \mathcal{S}_m$ and let $g : \mathbf{R}^m \to \mathbf{R}$ be a symmetric function. Then $g \circ \lambda$ is continuously differentiable at Y if and only if g is continuously differentiable at $\lambda(Y)$ and*

$$\nabla(g \circ \lambda)(Y) = Q \mathrm{diag}(\nabla f(\lambda(Y))) Q^T . \tag{2.5}$$

By using Lemma 3, we obtain

Proposition 4 *The augmented Lagrangian is continuously differentiable at x with*

$$\nabla_x F(x, U, \tau) = \nabla_x f(x) - (C_1(x) \bullet M_+(x, U, \tau), ..., C_n(x) \bullet M_+(x, U, \tau))^T . \tag{2.6}$$

Proof. Let $M_+^2(x, U, \tau) = Q \lambda(M_+^2(x, U, \tau)) Q^T$ be an orthogonal decomposition of $M_+^2(x, U, \tau)$. Then $\mathrm{Tr}[M_+^2(x, U, \tau)] = (g \circ \lambda)(M(x, U, \tau))$, where

$$g(v) = \sum_{i=1}^m [v_i]_+^2 \text{ and } v = \lambda(M(x, U, \tau)).$$

The function $g(v)$ is obviously symmetric. By Lemma 3, $\mathrm{Tr}[M_+^2(x,U,\tau)]$ is continuously differentiable at $M(x,U,\tau)$ with

$$\nabla_{M(x,U,\tau)}\mathrm{Tr}[M_+^2(x,U,\tau)] = Q[\nabla g(\lambda(M(x,U,\tau))]Q^T = 2M_+(x,U,\tau).$$

Now let e_i be the ith coordinate unit vector of $\mathbf{R}^n$ and let s be a scalar variable.

$$\begin{aligned}
\mathrm{Tr}[M_+^2(x+se_i,U,\tau)] &= \mathrm{Tr}\{[U-\tau C(x+se_i)]_+^2\} \\
&= \mathrm{Tr}\left\{[U-\tau C(x)-\tau s C_i(x)+o(s)]_+^2\right\} \\
&= \mathrm{Tr}\left\{[M(x,U,\tau)-\tau s C_i(x)+o(s)]_+^2\right\} \\
&= \mathrm{Tr}\left[M_+^2(x,U,\tau)\right] - \tau s C_i(x)\bullet\nabla_{M(x,U,\tau)}\mathrm{Tr}\left[M_+^2(x,U,\tau)\right]+o(s) \\
&= \mathrm{Tr}[M_+^2(x,U,\tau)] - 2\tau s C_i(x)\bullet M_+(x,U,\tau)+o(s).
\end{aligned}$$

Thus,

$$\nabla_{x_i}\mathrm{Tr}[M_+^2(x,U,\tau)] = -2\tau C_i(x)\bullet M_+(x,U,\tau) \tag{2.7}$$

and (2.6) is valid. □

Remark 5 *The formula 2.7 can be written as*

$$\nabla_{x_i}[M_+(x,U,\tau)\bullet M_+(x,U,\tau)] = \nabla_{M(x,U,\tau)}[M_+(x,U,\tau)\bullet M_+(x,U,\tau)]\bullet\nabla_{x_i}M(x,U,\tau),$$

actually showing a chain rule for differentiating $Tr[M_+^2(x,U,\tau)]$.

2.3 The second order differential of $F(x,U,\tau)$

Let $A,B\in\mathcal{S}^m$. Define a linear operator $\mathcal{L}_A:\mathcal{S}^m\mapsto\mathcal{S}^m$ as $\mathcal{L}_A(B)=AB+BA$.

Lemma 6 [20, Theorem 4.6] *The function* $[\cdot]_+:\mathcal{S}^m\mapsto\mathcal{S}^m$ *is F-differentiable at* $Y\in\mathcal{S}^m$ *if and only if* Y *is nonsingular. In this case,* $\nabla[Y]_+=\mathcal{L}_{|Y|}^{-1}\circ\mathcal{L}_{Y_+}$, *where* $\circ$ *denotes composition of operators and* $\mathcal{L}_{|Y|}^{-1}$ *is the inverse operator of* $\mathcal{L}_{|Y|}$. *That is* $Z=\mathcal{L}_{|Y|}^{-1}X \Longleftrightarrow X=\mathcal{L}_{|Y|}Z$.

Proposition 7 *If Assumptions (i)- (iii) hold, then* $M_+(x,U,\tau)$ *is differentiable with respect to* x, *at* (x^*,U^*) *for* $\tau>0$, *and*

$$\frac{\partial M_+(x^*,U^*,\tau)}{\partial x_j} = -\frac{\tau}{2}C_j(x^*)+\frac{1}{2}Y_j, \tag{2.8}$$

where

$$Y_j = -\tau\mathcal{L}_{|M(x^*,U^*,\tau)|}^{-1}\circ\mathcal{L}_{M(x^*,U^*,\tau)}(C_j(x^*)). \tag{2.9}$$

Proof. This is a direct implication of Lemma 6 since $M(x^*,U^*,\tau)$ is nonsingular by Lemma 2 and

$$\begin{aligned}
\mathcal{L}_{|M(x^*,U,\tau)|}^{-1}\circ\mathcal{L}_{M_+(x^*,U,\tau)} &= \mathcal{L}_{|M(x^*,U,\tau)|}^{-1}\circ\mathcal{L}_{(|M(x^*,U,\tau)|+M(x^*,U^*,\tau))/2} \\
&= \frac{1}{2}\left(I+\mathcal{L}_{|M(x^*,U^*,\tau)|}^{-1}\circ\mathcal{L}_{M(x^*,U^*,\tau)}\right),
\end{aligned}$$

where I is the identity operator from $\mathcal{S}^m$ to $\mathcal{S}^m$. □

Proposition 8 *If Assumptions (i)–(iv) hold, then $\nabla^2_{xx}F(x^*,U^*,\tau)$ is positive definite on $\mathcal{C}(x^*)$ when $\tau > 0$. Moreover, there is a positive scalar $\bar{\tau}$ such that $\nabla^2_{xx}F(x^*,U^*,\tau)$ is positive definite when $\tau \geq \bar{\tau}$.*

Proof. Since $M_+(x,U,\tau)$ is differentiable with respect to x at (x^*,U^*), $\nabla^2_{xx}F(x^*,U^*,\tau)$ is well defined and by 2.7

$$\frac{\partial^2 F(x^*,U^*,\tau)}{\partial x_j \partial x_i} = \frac{\partial^2 f(x^*)}{\partial x_j \partial x_i} - M_+(x^*,U^*,\tau) \bullet C_{ij}(x^*) - C_i(x^*) \bullet \frac{\partial M_+(x^*,U^*,\tau)}{\partial x_j}.$$

Thus $\nabla^2_{xx}F(x^*,U^*,\tau)$ may be expressed as

$$\nabla^2_{xx}F(x^*,U^*,\tau) = \nabla^2_{xx}L(x^*,U^*) + [-C_i(x^*) \bullet \partial M_+(x^*,U^*,\tau)/\partial x_j]^n_{i,j=1}.$$

From Proposition 7, we have

$$-C_i(x^*) \bullet \partial M_+(x^*,U^*,\tau)/\partial x_j = \frac{\tau}{2}\mathrm{vec}(C_i(x^*))^T \mathrm{vec}(C_j(x^*)) - \frac{1}{2}\mathrm{vec}(C_i(x^*))^T \mathrm{vec}(Y_j),$$

and we need to derive a formula for expressing $\mathrm{vec}(Y_j)$.
Since Y_j satisfies (2.9), we have from Lemma 2 that

$$\begin{aligned} & Q\,\mathrm{diag}\,(\Gamma_1,\tau\Gamma_2)Q^T Y_j + Y_j Q\,\mathrm{diag}\,(\Gamma_1,\tau\Gamma_2)Q^T \\ = \ & Q\,\mathrm{diag}\,(\Gamma_1,-\tau\Gamma_2)Q^T(-\tau C_j(x^*)) - \tau C_j(x^*)Q\,\mathrm{diag}\,(\Gamma_1,-\tau\Gamma_2)Q^T. \end{aligned}$$

Pre-multiplying and post-multiplying the above quality by Q^T and Q respectively, we obtain

$$\begin{aligned} & \mathrm{diag}\,(\Gamma_1,\tau\Gamma_2)Q^T Y_j Q + Q^T Y_j Q\,\mathrm{diag}\,(\Gamma_1,\tau\Gamma_2) \\ = \ & \mathrm{diag}\,(\Gamma_1,-\tau\Gamma_2)Q^T(-\tau C_j(x^*))Q + Q^T(-\tau C_j(x^*))Q\,\mathrm{diag}\,(\Gamma_1,-\tau\Gamma_2), \end{aligned}$$

namely,

$$\begin{aligned} & (I_m \otimes \mathrm{diag}\,(\Gamma_1,\tau\Gamma_2) + \mathrm{diag}\,(\Gamma_1,\tau\Gamma_2) \otimes I_m)\mathrm{vec}(Q^T Y_j Q) = \\ & (I_m \otimes \mathrm{diag}\,(\Gamma_1,-\tau\Gamma_2) + \mathrm{diag}\,(\Gamma_1,-\tau\Gamma_2) \otimes I_m)\mathrm{vec}(Q^T(-\tau C_j(x^*))Q). \end{aligned}$$

Thus

$$\mathrm{vec}(Q^T Y^j Q) = -\tau\mathcal{B}\mathrm{vec}(Q^T C_j(x^*)Q), \tag{2.10}$$

where

$$\begin{aligned} \mathcal{B} \quad & = (I_m \otimes \mathrm{diag}\,(\Gamma_1,\tau\Gamma_2) + \mathrm{diag}\,(\Gamma_1,\tau\Gamma_2) \otimes I_m)^{-1} \\ & \quad (I_m \otimes \mathrm{diag}\,(\Gamma_1,-\tau\Gamma_2) + \mathrm{diag}\,(\Gamma_1,-\tau\Gamma_2) \otimes I_m). \end{aligned}$$

Obviously, all eigenvalues of $\mathcal{B}$ are in the interval $[-1,1]$. By (2.10) we obtain

$$(Q^T \otimes Q^T)\mathrm{vec}(Y_j) = -\tau\mathcal{B}\mathrm{vec}(Q^T C_j(x^*)Q),$$

which implies

$$\mathrm{vec}(Y_j) = -\tau(Q \otimes Q)\mathcal{B}\mathrm{vec}(Q^T C_j(x^*)Q).$$

Therefore we have

$$\begin{aligned}
& -C_i(x^*) \bullet \partial M_+(x^*, U^*, \tau)/\partial x_j \\
= \; & \frac{\tau}{2}\mathrm{vec}(C_i(x^*))^T \mathrm{vec}(QQ^T C_j(x^*)QQ^T) - \frac{1}{2}\mathrm{vec}(C_i(x^*))^T \mathrm{vec}(Y_j) \\
= \; & \frac{\tau}{2}\mathrm{vec}(C_i(x^*))^T (Q \otimes Q)\mathrm{vec}(Q^T C_j(x^*)Q) - \\
& -\frac{1}{2}\mathrm{vec}(C_i(x^*))^T (Q \otimes Q)\mathcal{B}\mathrm{vec}(Q^T(-\tau C_j(x^*))Q) \\
= \; & \frac{\tau}{2}\mathrm{vec}(Q^T C_i(x^*)Q)^T [I_{m^2} + \mathcal{B}]\mathrm{vec}(Q^T C_j(x^*)Q).
\end{aligned}$$

Let

$$[\mathrm{vec}(Q^T C_1(x^*)Q), \cdots, \mathrm{vec}(Q^T C_n(x^*)Q)] = Z(x^*).$$

Then

$$[-C_i(x^*) \bullet \partial M_+(x^*, U^*, \tau)/\partial x_j]_{i,j=1}^n = \frac{\tau}{2} Z(x^*)^T [I_{m^2} + \mathcal{B}] Z(x^*) \succeq 0.$$

This, together with Assumption (iv), proves the positive definiteness of $\nabla_{xx}^2 F(x^*, U^*, \tau)$ on $\mathcal{C}(x^*)$ for any $\tau > 0$.

To prove the second part of the proposition, we split Q into two blocks $Q = [E \; F]$, where E is defined in Assumption (iv). Let γ_i be the ith diagonal entry of $\mathrm{diag}\,(\Gamma_1. - \tau\Gamma_2)$. Note that $Z(x^*)^T [I_{m^2} + \mathcal{B}] Z(x^*)$ can be expressed as $A_1(x^*) + A_2(x^*) + A_3(x^*)$, where

$$\begin{aligned}
A_1(x^*) &= (\partial(C(x^*)E)/\partial x)^T \begin{bmatrix} 2EE^T & & \\ & \ddots & \\ & & 2EE^T \end{bmatrix} \partial(C(x^*)E)/\partial x, \\
A_2(x^*) &= (\partial(C(x^*)E)/\partial x)^T \begin{bmatrix} \mathcal{D}_1 & & \\ & \ddots & \\ & & \mathcal{D}_{m-r} \end{bmatrix} \partial(C(x^*)E)/\partial x, \\
A_3(x^*) &= (\partial(C(x^*)F)/\partial x)^T \begin{bmatrix} 2EE^T - \mathcal{D}_{m-r+1} & & \\ & \ddots & \\ & & 2EE^T - \mathcal{D}_m \end{bmatrix} \partial(C(x^*)F)/\partial x
\end{aligned}$$

and

$$\mathcal{D}_i = \begin{cases} 2\gamma_i F(\tau\Gamma_2 + \gamma_i I_r)^{-1} F^T & \text{if} \quad 1 \le i \le m - r, \\[2ex] 2\gamma_i E(\tau\Gamma_1 + \gamma_i I_{m-r})^{-1} E^T & \text{if} \quad m - r + 1 \le i \le m. \end{cases}$$

As $A_2(x^*)$ and $A_3(x^*)$ are positive semidefinite, we have

$$\nabla_{xx}^2 F(x^*, U^*, t) \succeq \nabla_{xx}^2 L(x^*, U^*) + \frac{1}{2}\tau A_1(x^*).$$

Since $A_1(x^*)$ is positive definite outside $\mathcal{C}(x^*)$ and its eigenvalues are independent of τ, it follows from Lemma 1.25 of [2] and Assumption (iv) that there is a positive scalar $\bar{\tau}$ such that $\nabla_{xx}^2 F(x^*, U^*, \tau)$ is positive definite when $\tau \ge \bar{\tau}$. □

We now proceed to show that there exists a threshold value $\tau_0 > 0$ such that for all $\tau \geq \tau_0$ the minimum of $F(x, U^*, \tau)$ is a local minimum of (1.5). To this end, we need to show that under our assumptions the so-called quadratic growth condition holds. According to [3, Theorem 5.89], the quadratic growth condition holds under Assumption (i) if

$$\nabla^2_{xx} L(x^*, U^*) + H(x^*, U^*)$$

is positive definite over $\mathcal{C}(x^*)$, where

$$H(x, U) = 2(\frac{\partial C(x)}{\partial x})^T (U \otimes [C(x)]^\dagger) \frac{\partial C(x)}{\partial x},$$

i.e. its ijth entry is

$$H_{ij} = 2U \bullet C_i(x)[C(x)]^\dagger C_j(x), i, j = 1, \ldots, n, \tag{2.11}$$

where $[C(x)]^\dagger$ denotes the Moore-Penrose inverse of $C(x)$.

Lemma 9 $\nabla^2_{xx} L(x^*, U^*) + H(x^*, U^*)$ *is positive definite over* $\mathcal{C}(x^*)$.

Proof. Splitting the orthogonal matrix Q in the proof of Proposition 8 into $Q = [E, E']$ with E and E' correspond to Γ_1 and Γ_2, respectively. From the definition of H_{ij} (see (2.11)), we have

$$\begin{aligned} H_{ij} &= 2\text{vec}(C_i(x^*))^T (E\Gamma_1 E^T \otimes C(x^*)^\dagger) \text{vec}(C_j(x^*)) \\ &= 2\text{vec}(C_i(x^*))^T \text{vec}(C(x^*)^\dagger C_j(x^*) E\Gamma_1 E^T) \\ &= 2\text{vec}(C_i(x^*))^T (E\Gamma_1 \otimes C(x^*)^\dagger) \text{vec}(C_j(x^*)E) \\ &= 2\text{vec}(C(x^*)^\dagger C_i(x^*) E\Gamma_1)^T \text{vec}(C_j(x^*)E) \\ &= 2(\Gamma_1 \otimes C(x^*)^\dagger \text{vec}(C_i(x^*)E))^T \text{vec}(C_j(x^*)E) \\ &= 2\text{vec}(C_i(x^*)E)^T (\Gamma_1 \otimes C(x^*)^\dagger) \text{vec}(C_j(x^*)E) \\ &= \text{vec}(C_i(x^*)E)^T \mathcal{E} \text{vec}(C_j(x^*)E), \end{aligned}$$

where $\mathcal{E} = \text{diag}(\mathcal{E}_1, \ldots, \mathcal{E}_{m-r})$ with $\mathcal{E}_i = 2\gamma_i E' \Gamma_2^{-1} E'^T$ for $i = 1, \ldots, m-r$. Therefore we obtain

$$H(x^*, U^*) = [\partial(C(x^*)E)/\partial x]^T \mathcal{E} \partial(C(x^*)E)/\partial x \succeq 0.$$

In view of Assumption (iv), the lemma follows. □

Let $(x^*; U^*, \tau^*)$ be a saddle point of the augmented Lagrangian over $\mathbf{R}^n \times \mathcal{S}^m \times \mathbf{R}_+$. Namely, it satisfies

$$F(x, U^*, \tau) \geq F(x^*, U^*, \tau) \geq F(x^*, U, \tau) \;\; \text{for all } x \in \mathbf{R}^n, U \in \mathcal{S}^m. \tag{2.12}$$

(Define $F(x, U, 0) = L(x, U)$.) It is easy to show that x^* is a local minimum of (1.5) if (2.12) is satisfied. The following Theorem is a special case of Rockafellar [17, Corollary 6.1] that guarantees such a saddle point exists.

Theorem 10 *Suppose (1.5) satisfies the quadratic growth condition. Let x^* be a strict local minimum of (1.5). Assume that x^* satisfies Assumptions (i)- (iv) with U^* as the Lagrange multiplier. Then the global saddle point condition (2.12) holds for all τ not less than a threshold value $\tau_0 \geq \bar{\tau}$.*

Based on this theorem, various augmented Lagrange methods can be developed to compute (x^*, U^*), see for example [6, 9]. Our main interest now, however, is to analyze the local rate of convergence for the general process (1.3)-(1.4) in the case of semidefinite optimization. This is the task of the next section.

3 An augmented Lagrangian algorithm and its local convergence

We present an augmented Lagrangian algorithm and analyze its rate of local convergence.
Algorithm

Step 0 Let $0 < \tau_0 \leq \tau$. Set $U^0 \in S_m^{++}$ and set $k = 0$.

Step 1 Compute x^k so that

$$x^k = \operatorname{argmin}_{x \in \mathbf{R}^n} F(x, U^k, \tau)$$

Step 2 If $U^k C(x^k) = 0$, then stop (x^k is a KKT point).

Step 3 Update U^{k+1} by

$$U^{k+1} = U^k + \tau \nabla_U F(x^k, U^k, \tau) = [U^k - \tau C(x^k)]_+.$$

Step 4 Set $k := k + 1$ and go to step 1.

Theorem 11 *Let assumptions (i)-(iv) hold. Then there exist $\delta > 0$, $\epsilon > 0$, $\tau_0 > 0$ and τ_2 with $\tau_0 \leq \tau_1 < \tau_2$, such that for any $\tau \in [\tau_1, \tau_2]$ and for $U \in B(U^*, \delta)$, the following statements are true:*

(a) *There exists a unique vector $\hat{x} = \hat{x}(U, \tau)$ such that $\nabla_x F(\hat{x}, U, \tau) = 0$, and*

$$\hat{x} = \operatorname{argmin}_{x \in \mathbf{R}^n} \{F(x, U, \tau) | x \in B(x^*, \epsilon)\}. \tag{3.1}$$

(b) *Denote $\hat{M}_+ = M_+(\hat{x}(U, \tau), U, \tau)$. Assume that the inverse of*

$$A(U, \tau) = \begin{bmatrix} \nabla_{xx}^2 F(\hat{x}, \hat{M}_+, \tau) & -\partial C(\hat{x}(U, \tau))^T / \partial x \\ \\ -\frac{1}{2} \partial C(\hat{x}(U, \tau))^T / \partial x - \frac{1}{2} \mathcal{Z} & -\tau^{-1} I_{m^2 \times m^2} \end{bmatrix}$$

is bounded for U in a neighborhood of U^* and τ is sufficiently large, where $\mathcal{Z} = (vec(Z_1), \ldots, vec(Z_n))$ with $Z_i = L^{-1}_{|M(\hat{x}(U,\tau),U,\tau)|} \circ L_{M(\hat{x}(U,\tau),U,\tau)}(C_i(\hat{x}(U, \tau)))$. Then the estimates

$$\|\hat{x} - x^*\| \leq c\tau^{-1} \|U - U^*\|_F,$$

$$\|\hat{U} - U^*\|_F \leq c\tau^{-1} \|U - U^*\|_F$$

hold, where constant c is independent of τ and $\hat{U} = M_+(\hat{x}(U, \tau), U, \tau)$.

As pointed out by a referee, result (b) indicates that, as in the case of Augmented Lagrangian methods for standard nonlinear programs, the rate of local linear convergence can be freely adjusted by increasing τ — at the cost of possibly reducing the size of the domain of local convergence.

Proof. **(a)** The existence of $\hat{x}(U,\tau)$ is a direct implication of the classical implicit function theorem applied to $\nabla_x F(x,U,\tau)=0$ at (x^*,U^*,τ), just noting that $\nabla_x F(x^*,U^*,\tau)=\nabla_x L(x^*,U^*)=0$ because of Lemma 2, Proposition 4, and Assumption (ii) and noting that $\nabla^2_{xx}F(x^*,U^*,\tau)\succ 0$ for large τ due to Proposition 8. Now it follows from the implicit function theorem, in a neighborhood of (U^*,τ), one has $\nabla_x F(\hat{x},U,\tau)=0$. By continuity of $\nabla^2_{xx}F(x,U,\tau)$, one also has $\nabla^2_{xx}F(\hat{x},U,\tau)\succ 0$. Thus, (3.1) holds.
Now we prove **(b)**. Let us differentiate (3.3) and (3.4) with respect to U and τ^{-1}, yielding

$$\begin{bmatrix} \nabla^2_{xx}F(\hat{x},\hat{M}_+,\tau) & -\partial C(\hat{x})^T/\partial x \\ -\frac{1}{2}\partial C(\hat{x})^T/\partial x-\frac{1}{2}\mathcal{Z} & -\tau^{-1}I_{m^2\times m^2}\end{bmatrix}\begin{bmatrix} \nabla^T_{\mathrm{vec}(U)}\hat{x}(U,\tau) & \frac{\partial \hat{x}(U,\tau)}{\partial(\tau^{-1})} \\ \nabla^T_{\mathrm{vec}(U)}\mathrm{vec}\left(\hat{M}_+\right) & \frac{\partial \mathrm{vec}\hat{M}_+}{\partial(\tau^{-1})}\end{bmatrix}+$$

$$+\begin{bmatrix} 0 & 0 \\ \tau^{-1}\nabla_{\mathrm{vec}(U)}\mathrm{vec}\left(M_+\right)|_{x=\hat{x}(U,\tau)} & -\hat{M}_+ +\frac{\partial}{\partial(\tau^{-1})}\mathrm{vec}[\tau^{-1}M_+]|_{x=\hat{x}(U,\tau)}\end{bmatrix}=0,$$

where $\mathcal{Z}=(\mathrm{vec}(Z_1),\ldots,\mathrm{vec}(Z_n))$ with $Z_i=L^{-1}_{|M(\hat{x}(U,\tau),U,\tau)|}\circ L_{M(\hat{x}(U,\tau),U,\tau)}(C_i(\hat{x}(U,\tau)))$. From Proposition 7, we choose δ sufficiently small such that $U-\tau C(\hat{x}(U,\tau))$ being nonsingular. Then we obtain

$$\nabla_{\mathrm{vec}(U)}\mathrm{vec}\left(M_+\right)|_{x=\hat{x}(U,\tau)}=I_{m^2\times m^2}/2+\mathcal{E}/2,$$

where $\mathcal{E}=\{\mathrm{vec}(\mathcal{E}_{i,j})\,|\,1\le i,j\le m\}$ satisfies

$$\sqrt{(U-\tau C(\hat{x}(U,\tau)))^2}\mathcal{E}_{i,j}+\mathcal{E}_{i,j}\sqrt{(U-\tau C(\hat{x}(U,\tau)))^2}=$$
$$(U-\tau C(\hat{x}(U,\tau)))e_ie_j^T+e_ie_j^T(U-\tau C(\hat{x}(U,\tau))).$$

Let $G(U,\tau)=\partial/\partial(\tau^{-1})\,\mathrm{vec}[\tau^{-1}M_+(x,U,\tau)]|_{x=\hat{x}(U,\tau)}$. Then

$$\begin{aligned} G(U,\tau)= \quad & 1/2\mathrm{vec}(U)+ \\ & +1/2\left(I_m\otimes\sqrt{(\tau^{-1}U-C(\hat{x}(U,\tau)))^2}+\sqrt{(\tau^{-1}U-C(\hat{x}(U,\tau)))^2}\otimes I_m\right)^{-1}\times \\ & \times\left(I_m\otimes(\tau^{-1}U-C(\hat{x}(U,\tau)))+(\tau^{-1}U-C(\hat{x}(U,\tau)))\otimes I_m\right)\mathrm{vec}(U).\end{aligned}$$

Now we estimate the bound of $G(U,\tau)$ under the condition that $\tau^{-1}U-C(\hat{x}(U,\tau))$ has an inverse and its inverse is bounded. Let $\Phi(Y)=|Y|$.

$$G(U,\tau)=\mathrm{vec}(U)/2+\mathrm{vec}\,\Phi'(Y,U)/2,$$

with $Y = \tau^{-1}U - C(\hat{x}(U,\tau))$. In view of Lemma 6,

$$\begin{aligned}
G(U,\tau) &= \mathrm{vec}(U)/2 + (1/2)(I \otimes |Y| + |Y| \otimes I)^{-1}(I \otimes Y + Y \otimes I)\mathrm{vec}(U) \\
&= \mathrm{vec}(U) + (1/2)I \otimes |Y| + |Y| \otimes I)^{-1}(I \otimes (Y - |Y|) + (Y - |Y|) \otimes I)\mathrm{vec}(U) \\
&= \mathrm{vec}(U) - (I \otimes |Y| + |Y| \otimes I)^{-1}(I \otimes [-Y]_+ + [-Y]_+ \otimes I)\mathrm{vec}(U) \\
&= \mathrm{vec}(U) - (I \otimes |Y| + |Y| \otimes I)^{-1}\mathrm{vec}([-Y]_+U + U[-Y]_+),
\end{aligned}$$

and it follows from $[-Y^*]_+ = C(x^*)$ and $C(x^*)U^* = U^*C(x^*) = 0$ that

$$\begin{aligned}
& \|\mathrm{vec}([-Y]_+U + U[-Y]_+)\| \\
= & \Big\| ([-Y]_+ - [-Y^*]_+)(U - U^*) + (U - U^*)([-Y]_+ - [-Y^*]_+) + \\
& \quad +[-Y^*]_+(U - U^*) + (U - U^*)[-Y^*]_+ + \\
& \quad + ([-Y]_+ - [-Y^*]_+)U^* + U^*([-Y]_+ - [-Y^*]_+) \Big\| \\
\leq & \ 2\,\|[-Y]_+ - [-Y^*]_+\|_F\,\|U - U^*\|_F + \\
& \quad +2\,\|[-Y^*]_+\|_F\,\|U - U^*\|_F + 2\|U^*\|_F\,\|[-Y]_+ - [-Y^*]_+\|_F \\
\leq & \ 2\,(\|U^*\|_F + \|U - U^*\|_F)\,\|[-Y]_+ - [-Y^*]_+\|_F + \\
& \qquad +2\|C(x^*)\|_F\|U - U^*\|_F \\
\leq & \ 2\,(\|U^*\|_F + \|U - U^*\|_F)\,\|[-Y] - [-Y^*]\|_F + 2\|C(x^*)\|_F\|U - U^*\|_F.
\end{aligned}$$

Since C is Lipschitz continuous at x^*, we have $\|C(\hat{x}(U,\tau)) - C(x^*)\| \leq \varsigma_1\|\hat{x}(U,\tau) - x^*\|$ and

$$\begin{aligned}
\|Y - Y^*\|_F &= \|(\tau^{-1}U - C(\hat{x}(U,\tau))) - (\tau^{-1}U^* - C(x^*))\|_F \\
&\leq \tau^{-1}\|U - U^*\|_F + \varsigma_1\|\hat{x}(U,\tau) - x^*\|.
\end{aligned}$$

Therefore,

$$\|\mathrm{vec}([-Y]_+U + U[-Y]_+)\| \leq \varsigma_2\tau^{-1}\|U - U^*\|_F + \varsigma_3\|U - U^*\|_F + \varsigma_4\|\hat{x}(U,\tau) - x^*\|,$$

where $\varsigma_2 = 2(\|U^*\|_F + \delta)$, $\varsigma_3 = 2\|C(x^*)\|_F$ and $\varsigma_4 = \varsigma_1\varsigma_2$ and

$$\begin{aligned}
& \|\mathrm{vec}[\hat{M}_+(U,\tau) - G(U,\tau)]\| \\
= & \ \|\hat{M}_+(U,\tau) - G(U,\tau)\|_F \\
\leq & \ \|\hat{M}_+(U,\tau) - U\|_F + \beta_0\|\mathrm{vec}([-Y]_+U + U[-Y]_+)\| \\
\leq & \ \|\hat{M}_+(U,\tau) - U\|_F + \beta_0[\varsigma_2\tau^{-1}\|U - U^*\|_F + \varsigma_3\|U - U^*\|_F + \varsigma_4\|\hat{x}(U,\tau) - x^*\|],
\end{aligned}$$

where β_0 is a bound of $\|(I \otimes |Y| + |Y| \otimes I)^{-1}\|$ for U in a neighborhood of U^* and and $\tau \geq \tau_0$ sufficiently large.
Similar to the proof of the positive definiteness of $\nabla^2_{xx}F(x^*, U^*, \tau)$ in Proposition 8, we can prove that the Schur complement of $-\tau^{-1}I_{m^2 \times m^2}$ in $A(U,\tau)$, namely

$$\nabla^2_x F(\hat{x}(U,\tau), \hat{M}_+(U,\tau)) + \frac{1}{2}\tau\partial C(\hat{x}(U,\tau))^T/\partial x(\partial C(\hat{x}(U,\tau))^T/\partial x + \mathcal{Z}),$$

is positive definite. Thus, $A(U,\tau)$ is nonsingular for U in a neighborhood of U^* and $\tau \geq \tau_0$ sufficiently large. Let (from the assumption in (b)) μ be a bound of $\|A(U,\tau)^{-1}\|$ for

$(U,\tau) \in \{U' \mid \|U - U^*\| \le \delta\} \times [\bar\tau(1+\delta\bar\tau)^{-1}, \bar\tau(1-\delta\bar\tau)^{-1}]$ (we can choose $\bar\tau$ large enough and $\delta > 0$ small enough if necessary). We have for all (U,τ), such that $\|U - U^*\| \le \delta$ and $\tau \in [\bar\tau(1+\delta\bar\tau)^{-1}, \bar\tau(1-\delta\bar\tau)^{-1}]$,

$$\begin{aligned}
\begin{bmatrix} \hat x(U,\tau) - x^* \\ \\ \mathrm{vec}(\hat M_+(U,\tau) - U^*) \end{bmatrix} &= \begin{bmatrix} \hat x(U,\tau) - \hat x(U^*, 2\tau) \\ \\ \mathrm{vec}(\hat M_+(U,\tau) - \hat M_+(U^*, 2\tau)) \end{bmatrix} \\
= \quad & \int_0^1 \begin{bmatrix} \nabla^T_{\mathrm{vec}(U)} \hat x(U(\alpha),\tau(\alpha)) & \partial \hat x(U(\alpha),\tau(\alpha))/\partial\tau^{-1} \\ \nabla^T_{\mathrm{vec}(U)} \mathrm{vec}\hat M_+(U(\alpha),\tau(\alpha)) & \partial \mathrm{vec}\hat M_+(U(\alpha),\tau(\alpha))/\partial\tau^{-1} \end{bmatrix} \mathrm{d}\alpha \begin{bmatrix} \mathrm{vec}(U-U^*) \\ \tau^{-1}/2 \end{bmatrix} \\
= \quad & \int_0^1 A(U(\alpha),\tau(\alpha))^{-1} \begin{bmatrix} 0 & 0 \\ \frac{1}{2\tau}(I + \mathcal{E}(U(\alpha),\tau(\alpha))) & \hat M_+(U(\alpha),\tau(\alpha)) - G(U(\alpha),\tau(\alpha)) \end{bmatrix} \mathrm{d}\alpha \begin{bmatrix} \mathrm{vec}(U-U^*) \\ \tau^{-1}/2 \end{bmatrix}
\end{aligned}$$

where $U(\alpha) = U^* + \alpha(U - U^*)$ and $\tau(\alpha) = 2(1+\alpha)^{-1}\tau$. Therefore,

$$\begin{aligned}
& (\|\hat x(U,\tau) - x^*\|^2 + \|\hat M_+(U,\tau) - U^*\|_F^2)^{1/2} \\
& \le \mu[\beta_1\tau^{-1}\|U - U^*\|_F + (\tau^{-1}/2) \max_{0\le\alpha\le 1} \|\hat M_+(U(\alpha),\tau(\alpha)) - G(U(\alpha),\tau(\alpha))\|_F] \\
& \le \mu[\beta_1\tau^{-1}\|U - U^*\|_F + (\tau^{-1}/2) \max_{0\le\alpha\le 1} (\|\hat M_+(U(\alpha),\tau(\alpha)) - U^*\|_F + \\
& + \|U(\alpha) - U^*\|_F + \beta_0(\varsigma_2\tau^{-1}\|U(\alpha) - U^*\|_F + \varsigma_3\|U(\alpha) - U^*\|_F + \varsigma_4\|\hat x(U(\alpha),\tau(\alpha)) - x^*\|))] \\
& \le \varrho\tau^{-1}\|U - U^*\|_F + (1/2)\tau^{-1} \max_{0\le\alpha\le 1} [\|\hat M_+(U(\alpha),\tau(\alpha)) - U^*\|_F + \varsigma_4\|\hat x(U(\alpha),\tau(\alpha)) - x^*\|],
\end{aligned}$$

where β_1 is a bound of $(1/2)(I + \mathcal{E}(U(\alpha),\tau(\alpha)))$ and $\varrho = \mu\beta_1 + 1/2 + (1/2)\beta_0\varsigma_3 + (1/2)\beta_0\varsigma_2{\tau_0}^{-1}$. Since

$$\begin{aligned}
& \max\{\|\hat x(U(\alpha),\tau(\alpha)) - x^*\|_F, \|\hat M_+(U(\alpha),\tau(\alpha)) - U^*\|_F\} \\
\le \quad & \varrho\tau^{-1}\|U - U^*\|_F + (1/2)\tau^{-1} \max_{0\le\gamma\le 1} [\|\hat M_+(U(\alpha\gamma),\tau(\alpha\gamma)) - U^*\|_F + \\
& + \varsigma_4\|\hat x(U(\alpha\gamma),\tau(\alpha\gamma)) - x^*\|] \\
\le \quad & \varrho\tau^{-1}\|U - U^*\|_F + \max_{0\le\gamma\le 1} \left[\|\hat M_+(U(\gamma),\tau(\gamma)) - U^*\|_F + \varsigma_4\|\hat x(U(\gamma),\tau(\gamma)) - x^*\|\right],
\end{aligned}$$

we have that

$$\max_{0\le\alpha\le 1} \left[\|\hat M_+(U(\alpha),\tau(\alpha)) - U^*\|_F + \varsigma_4\|\hat x(U(\alpha),\tau(\alpha)) - x^*\|\right] \le \frac{(1+\varsigma_4)\varrho\tau^{-1}\|U - U^*\|_F}{1 - (1+\varsigma_4)\tau^{-1}/2},$$

which leads to

$$\left(\|\hat x(U,\tau) - x^*\|^2 + \|\hat M_+(U,\tau) - U^*\|_F^2\right)^{1/2} \le \tau^{-1}\|U - U^*\|_F \frac{(1+\varsigma_4)\varrho\tau^{-1}}{2 - (1+\varsigma_4)\tau^{-1}}. \tag{3.5}$$

We assume, without loss of generality, that $(1+\varsigma_4){\tau_0}^{-1} \le 1$, then

$$\frac{(1+\varsigma_4)\varrho\tau^{-1}}{2 - (1+\varsigma_4)\tau^{-1}} \le (1+\varsigma_4)\varrho{\tau_0}^{-1}$$

for $\tau \in [\tau_1, \tau_2]$ with $\tau_1 = \bar\tau(1+\delta\bar\tau)^{-1} > \tau_0$ and $\tau_2 = \bar\tau(1-\delta\bar\tau)^{-1}$. Let $c = (1+\varsigma_4)\varrho{\tau_0}^{-1}$, then the inequalities of **(b)** come from (3.5) with $\hat x = \hat x(U,\tau)$ and $\hat U = \hat M_+(U,\tau)$. The proof of **(b)** is therefore completed. □

References

[1] Alizadeh, F., (1995) *Interior point methods in semidefinite programming with application to combinatorial optimization*, SIAM J. Optim., 5:13-51.

[2] Bertsekas, D. P., (1982) *Optimization and Multiplier Methods*, Academic Press.

[3] J.F. Bonnans and A. Shapiro, (2000) *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York.

[4] Chen, X., Qi, H. and Tseng, P. , (2003) *Analysis of nonsmooth symmetric-matrix functions with applications to semidefinite complementarity problems*, SIAM J. Optim., 13:960-985.

[5] Chen, X. and Tseng, P., (2003) *Non-interior continuation methods for solving semidefinite complementarity problems*, Mathematical Programming, 95:431-474.

[6] Doljansky, M., Teboulle, M., (1999) *An interior proximal algorithm and the exponential multiplier method for semidefinite programming*, SIAM J. Optim., 9:1-13.

[7] Forsgren, A., (2000) *Optimality conditions for nonconvex semidefinite programming*, Math. Program. 88 : 105-128.

[8] Horn, R.A. and Johnson, C.R., (1991) *Topics in Matrix Analysis*, Cambridge University Press, Cambridge.

[9] Iusem, A. and Otero, R.G., (2001) *Augmented Lagrangian methods for cone-constrained convex optimization in Banach spaces*, submitted to Journal of Nonlinear and Convex Analysis.

[10] Lewis, A.S., (1996) *Derivatives of spectral functions,* Mathematics of Operations Research, 21:576–588.

[11] Lewis, A.S., (1999) *Nonsmooth analysis of eigenvalues,* Mathematical Programming, 84:1-24.

[12] Lewis, A.S. and Overton, M.L., (1996) *Eigenvalue optimization,* Acta Numerica, 5:149-190.

[13] Lewis, A.S. and Sendov, H.S., (2001) *Twice differentiable spectral functions,* SIAM Journal on Matrix Analysis, 23:368–386.

[14] Leibfritz, L., (2001) *An LMI based algorithm for designing suboptimal static H_2/H_∞ feedback controllers*, SIAM J. Optim. Control, 39:1711 - 1735.

[15] Zhang, L. W. and Liu, Y.J., (2003) *Convergence analysis of a nonlinear Lagrange algorithm for nonlinear programming with inequality constraints*, Journal of Applied Mathematics & Computing, 13:1-10.

[16] Pang, J.-S., Sun, D. and Sun, J., (2003) *Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems*, Mathematics of Operations Research, 28:39-63.

[17] Rockafellar, R.T., (1974) *Augmented Lagrange multiplier functions and duality in nonconvex programming*, SIAM J. Control, 12:268-285.

[18] Shapiro, A., (1997) *First and second order analysis of nonlinear semidefinite programs*, Math. Program.,77:301-320.

[19] Shapiro, A. and Sun, J., (2004) *Some properties of the augmented Lagrangian in cone constrained optimization*, Mathematics of Operations Research, 29:479-491.

[20] Sun, D. and Sun, J., (2002) *Semismooth matrix valued functions*, Mathematics of Operations Research, 27:150-169.

[21] Sun, J., Sun, D. and Qi, L., (2004) *A squared smoothing Newton method for nonsmooth matrix equations and its applications in semidefinite optimization problems*, SIAM J. Optim., 14:783-806.

[22] Vandenberghe,L., Boyd,S., (1996) *Semidefinite programming*, SIAM Rev. ,38:49-95.

[23] Wolkowicz, H., Saigal, R., Vandenberghe L., (2000) *Handbook of Semidefinite Programming: Theory, Algorithms,and Applications*, Kluwer Academic Publishers.

[24] Zhang, L.W. and He, S.X., (2000) *The convergence of a dual algorithm for nonlinear programming*, Korean J. Comput.& Appl.Math., 7(3):478-506.