Analysis of the primal-dual central path for nonlinear semidefinite optimization without the nondegeneracy condition

Takayuki Okuno
Faculty of Science and Technology Department of Science and Technology, Seikei University, 3-3-1, Kita, Kichijoji, Musashino, Tokyo, 180-8633, Japan
Center for Advanced Intelligence Project, RIKEN, Nihonbashi 1-chome Mitsui Building, 15th floor, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan, takayuki-okuno@st.seikei.ac.jp,

We study properties of the central path underlying a nonlinear semidefinite optimization problem, called NSDP for short. The latest radical work on this topic was contributed by Yamashita and Yabe (2012): they proved that the Jacobian of a certain equation-system derived from the Karush-Kuhn-Tucker (KKT) conditions of the NSDP is nonsingular at a KKT point under the second-order sufficient condition (SOSC), the strict complementarity condition (SC), and the nondegeneracy condition (NC). This yields uniqueness and existence of the central path through the implicit function theorem.

In this paper, we consider the following three assumptions on a KKT point: the strong SOSC, the SC, and the Mangasarian-Fromovitz constraint qualification. Under the absence of the NC, the Lagrange multiplier set is not necessarily a singleton and the nonsingularity of the above-mentioned Jacobian is no longer valid. Nonetheless, we establish that the central path exists uniquely, and moreover prove that the dual component of the path converges to the so-called analytic center of the Lagrange multiplier set. As another notable result, we clarify a region around the central path where Newton's equations relevant to primal-dual interior point methods are uniquely solvable.

Key words: nonlinear semidefinite optimization, primal-dual interior-point method, central path, nondegeneracy condition

---

1. Introduction

We consider the following nonlinear semidefinite optimization problem:

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & G(x) \in \mathbb{S}^m_+, \\ & h(x) = 0, \end{array} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{S}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^s$ are twice continuously differentiable functions. Moreover, $\mathbb{S}^m$ denotes the set of real $m \times m$ symmetric matrices and $\mathbb{S}^m_{++}$ (resp. $\mathbb{S}^m_+$) stands for the set of $m \times m$ real symmetric positive definite (resp. semidefinite) matrices. Throughout the paper, we often refer to problem (1.1) as NSDP. NSDP (1.1) contains a wide class of optimization problems. Indeed, when all the functions are affine with respect to x, it reduces to a linear semidefinite optimization problem (Vandenberghe and Boyd [56], Wolkowicz et al. [57]). When the function G is of the diagonal matrix form, it is regarded as a conventional nonlinear optimization problem (Mangasarian [37], Luenberger and Ye [35]). Moreover, it contains nonlinear second-order cone optimization problems (Kato and Fukushima [26], Bonnans and Ramírez [6]) by restricting the form of G appropriately.

The recent advance of researches on the NSDP is remarkable. Abundant practical applications of the NSDP can be found in a wide variety of fields, for example, structural optimization (Kočvara and Stingl [28], Thore et al. [54], Takezawa et al. [52], Thore [53]), control (Scherer [46], Kočvara et al. [27], Hoi et al. [18], Leibfritz and Volkwein [33]), statics (Qi and Sun [45]), finance (Konno et al. [30], Leibfritz and Maruhn [31]), positive semidefinite factorization (Vandaele et al. [55]), and so on. Elegant theoretical results on optimality conditions for the NSDP have been also developed. For example, the Karush-Kuhn-Tucker (KKT) conditions and the second-order conditions for the NSDP were studied in detail by Shapiro

[47] and Forsgren [10]. Further examples are: the strong second-order conditions by Sun [49], sequential optimality conditions by Andreani et al. [2], the local duality by Qi [44], and the optimality conditions via squared slack variables by Lourenço et al. [34]. Along with such theoretical results, various algorithms have been proposed for solving the NSDP, for example, augmented Lagrangian methods (Kočvara and Stingl [28], Sun et al. [51, 50], Andreani et al. [2, 1], Fukuda and Lourenço [12], Huang et al. [20], Wu et al. [59]), sequential linear semidefinite optimization methods (Kanzow et al. [24]), sequential quadratic semidefinite optimization methods (Correa and Ramirez C [8], Freund et al. [11], Zhao and Chen [67, 68], Yamakawa and Okuno [60]), sequential quadratically constrained quadratic semidefinite optimization methods (Auslender [4]), exact penalty methods (Auslender [5]), interior point-type methods (Arahata et al. [3], Jarre [21], Kato et al. [25], Leibfritz and Mostafa [32], Okuno and Fukushima [43, 42], Okuno [41], Yamashita and Yabe [63], Yamashita et al. [64, 65], Yamakawa and Yamashita [62, 61]), homotopy method (Yang and Yu [66]), and so forth.

In this paper, we study properties of the central path for the NSDP. The conventional central path is a path-like set formed by stationary points of the log-barrier penalized problem, and is a key concept of interior-point methods, abbreviated as IPMs, for solving a wide class of optimization problems including the NSDP. Many IPMs share the strategy of approaching a KKT point by following the central path approximately. Since the geometry of the central path is related to the performance of IPMs, it has been well studied under various settings. For example, Megiddo [38] presented an early work in this line for linear optimization or linear programming. Kojima et al. [29] and Monteiro and Tsuchiya [39] studied the central path for monotone complementarity problems under the absence of strict complementarity condition. Monteiro and Zou [40] worked with the existence of the central path for convex optimization problems. Wright and Orban [58] considered nonlinear optimization problems and analyzed the properties of the central path under the absence of linear independence constraint qualification.

We briefly review the history of the central path of semidefinite optimization problems (SDPs). Concerning linear SDPs, Luo et al. [36] showed that the (primal-dual) central path converges to the analytic center under the presence of the strict complementarity condition. Sturm and Zhang [48] further proved that the derivative of the central path is convergent. Halická et al. [16] proved that the central path is convergent regardless of the strict complementarity, by means of the curve selection lemma from algebraic geometry, although it can fail to converge to the analytic center in the absence of the strict complementarity. Halická [15] established that the central path is analytic including the boundary. See also other researches by Goldfarb and Scheinberg [13], Halická et al. [17], Kakihara et al. [22, 23], da Cruz Neto et al. [9], and so forth. More generally, Graña Drummond and Peterzil [14] worked with the existence and convergence of the central path of convex smooth SDP by assuming that the functions organizing the problem are analytic. Meanwhile, beside these studies concerning linear and convex SDPs, those for the general NSDP (1.1) are very scarce.

The latest radical work for NSDP (1.1) along this research-topic was presented by Yamashita and Yabe [63]. The authors analyzed the local convergence property of the primal-dual IPM, called PDIPM for short, that was proposed in their another article (Yamashita et al. [64]). This PDIPM is explained briefly as follows.: In the algorithm, the barrier KKT (BKKT) conditions are derived by perturbing the KKT conditions, and the degree of perturbation is controlled by the so-called barrier parameter. See Section 2.3 for the precise definition of the BKKT conditions. The PDIPM approaches a KKT point by generating a sequence of approximate BKKT points with driving the barrier parameter to zero. To compute a BKKT point, the Newton method combined with scaling techniques is applied to an equation-system equivalent to the BKKT conditions. In [63], Yamashita and Yabe proved that the Jacobian of this equation-system is nonsingular at a KKT point under the following three conditions: the strict complementarity condition (SC), the second-order sufficient condition (SOSC), and the nondegeneracy condition (NC). Along with the classical implicit function theorem, this fact yields that there exists a unique smooth path, i.e., a central path, passing through the focused KKT point, and this path is formed by BKKT points.

Our contribution

Main contribution of this paper is summarized as follows:

1. We prove that there exists a smooth central path under the SC, the strong SOSC, and the Mangasarian-Fromovitz constraint qualification (MFCQ). We also prove that the central path converges to the analytic center that is defined afterwards. Since the NC is not assumed therein, the Lagrange multiplier set is compact and convex, but not necessarily a singleton, although the KKT point is a strict local optimum due to the strong SOSC. This implies that the Jacobian described above is not always nonsingular, and thus the implicit function theorem is not applicable straightforwardly unlike [63].
2. Under the same conditions, we give a region around the central path where the Newton equation is solvable uniquely when applying the PDIPM.

In literature on SDPs, the analysis exploits the fact that the functions are analytic and, as a result, so is the underlying central path. However, this methodology is no longer applicable in our setting since the functions of the NSDP are not assumed to be analytic. The manner of the conducted analysis is motivated from Wright and Orban [58] for nonlinear optimization, but ours is more complicated because the SOSC of the NSDP involves difficulty arising from the so-called sigma term. Furthermore, we deal with the nonlinear equality constraints together, whereas [58] does not.

Notations and terminologies

Throughout the paper, we use the following notations as necessary: We denote the identity matrix in $\mathbb{R}^{m\times m}$ by I. For $A \in \mathbb{R}^{m\times m}$, we define $\mathrm{Sym}(A) := (A + A^\top)/2$ and $\|A\|_{\mathrm{F}} := \sqrt{\mathrm{trace}(A^\top A)}$. For $X, Y \in \mathbb{S}^m$, we define the inner product $X \bullet Y$ by $X \bullet Y := \mathrm{trace}(XY)$. We also define the linear operator $\mathcal{L}_X : \mathbb{S}^m \to \mathbb{S}^m$ by

$$\mathcal{L}_X(Y) := XY + YX.$$

Denote the smallest eigenvalue of $X \in \mathbb{S}^m$ by $\lambda_{\min}(X)$. For $X \in \mathbb{S}^m_+$ and $r > 0$, we denote by $X^{\frac{1}{r}}$ the unique solution $U \in \mathbb{S}^m$ of $U^r = X$. For a function $g : \mathbb{R}^n \to \mathbb{R}$, we denote by $\nabla g(x)$ or $\nabla_x g(x)$ the gradient of g, namely, $\nabla g(x) := (\frac{\partial g(x)}{\partial x_1}, \ldots, \frac{\partial g(x)}{\partial x_n})^\top \in \mathbb{R}^n$ and, also denote by $\nabla^2_{xx} g(x)$ the hessian of g, namely, $\nabla^2_{xx} g(x) = (\frac{\partial^2 g(x)}{\partial x_i \partial x_j})_{1\le i,j\le n} \in \mathbb{R}^{n\times n}$. For $\{A_k\}$ in a formed vector space with norm $\|\cdot\|$ and $\{b_k\} \subseteq \mathbb{R}$, we write $A_k = \mathrm{O}(b_k)$ if there exists some $M > 0$ such that $\|A_k\| \le M|b_k|$ for all k sufficiently large, and write $A_k = \mathrm{o}(b_k)$ if there exists some negative sequence $\{\alpha_k\} \subseteq \mathbb{R}$ such that $\lim_{k\to\infty} \alpha_k = 0$ and $\|A_k\| \le \alpha_k |b_k|$ for all k sufficiently large. We also say $A_k = \Theta(b_k)$ if there exist $M_1, M_2 > 0$ such that $M_1|b_k| \le \|A_k\| \le M_2|b_k|$ for all k sufficiently large.

We also denote $\mathbb{R}_{++} := \{a \in \mathbb{R} \mid a > 0\}$ and

$$\mathcal{W} := \mathbb{R}^n \times \mathbb{S}^m \times \mathbb{R}^s, \ \mathcal{W}_{++} := \{w \in \mathcal{W} \mid G(x) \in \mathbb{S}^m_{++}, Y \in \mathbb{S}^m_{++}\}.$$

Additionally, let $\mathcal{W}_+$ be the set obtained by replacing $\mathbb{S}^m_{++}$ with $\mathbb{S}^m_+$ in $\mathcal{W}_{++}$. For $w := (x, Y, z) \in \mathcal{W}$, $\|w\| := \sqrt{\|x\|_2^2 + \|Y\|_{\mathrm{F}}^2 + \|z\|_2^2}$, where $\|\cdot\|_2$ denotes the Euclidean norm.

Lastly, relevant to the function G in NSDP (1.1), we define the following notations. For $i = 1, 2, \ldots, n$, we write

$$\mathcal{G}_i(x) := \frac{\partial G(x)}{\partial x_i}.$$

For any $x, d \in \mathbb{R}^n$ and $Y \in \mathbb{S}^m$, we write

$$\Delta G(x; d) := \sum_{i=1}^n d_i \mathcal{G}_i(x) \in \mathbb{S}^m, \ \mathcal{J}G(x)^* Y := [\mathcal{G}_1 \bullet Y, \mathcal{G}_2 \bullet Y, \ldots, \mathcal{G}_n \bullet Y]^\top \in \mathbb{R}^n.$$

Some more notations and symbols will be introduced for the main analysis. See the paragraph *Additional notations and symbols used hereafter* at the end of subsection 3.1.

Organization of the paper

The rest of the paper is organized as follows: In section 2, we review some important concepts related to the NSDP such as the KKT conditions. In section 3, the main analysis is presented. In section 4, we conclude this paper with some remarks.

2. Preliminaries

2.1. KKT conditions for NSDP

We introduce the KKT conditions for NSDP (1.1).

Definition 1. We say that the the Karush-Kuhn-Tucker (KKT) conditions for NSDP (1.1) hold at $x \in \mathbb{R}^n$ if there exist a Lagrange multiplier matrix $Y \in \mathbb{S}^m$ and vector $z \in \mathbb{R}^s$ such that

$$\nabla_x L(w) = \nabla f(x) - \mathcal{J}G(x)^* Y + \nabla h(x) z = 0, \tag{2.1}$$

$$G(x) \bullet Y = 0,\ G(x) \in \mathbb{S}^m_+,\ Y \in \mathbb{S}^m_+, \tag{2.2}$$

$$h(x) = 0, \tag{2.3}$$

where $w := (x, Y, z) \in \mathcal{W}$ and $L : \mathcal{W} \to \mathbb{R}$ denotes the Lagrange function for the NSDP, that is,

$$L(w) := f(x) - G(x) \bullet Y + h(x)^\top z \tag{2.4}$$

for any $w \in \mathcal{W}$. Particularly, we call a triplet $w = (x, Y, z)$ satisfying the KKT conditions a KKT triplet of NSDP (1.1), and also call x a KKT point of the NSDP. Moreover, given a KKT point x, we denote by $\Lambda(x)$ the set of Lagrange multiplier pairs (Y, z) satisfying the KKT conditions at x, namely,

$$\Lambda(x) := \{(Y, z) \text{ satisfying (2.1)-(2.3)}\}.$$

Below, we define the Mangasarian-Fromovitz constraint qualification (MFCQ), under which the KKT conditions are ensured to be necessary optimality conditions for the NSDP.

Definition 2. ([7, Definition 2.8.6]) Let $x \in \mathbb{R}^n$ be a feasible point of NSDP (1.1). We say that the Mangasarian-Fromovitz constraint qualification (MFCQ) holds at x if $\nabla h(x)$ is of full column rank and there exists a vector $d \in \mathbb{R}^n$ such that $G(x) + \Delta G(x; d) \in \mathbb{S}^m_{++}$ and $\nabla h(x)^\top d = 0$.

Remark 1. Let $x \in \mathbb{R}^n$ be a local optimum of NSDP (1.1). Under the MFCQ, the KKT conditions hold at x, thus $\Lambda(x) \neq \emptyset$. In particular, the MFCQ implies that $\Lambda(x)$ is convex and also compact. Conversely, when f is convex, h is affine, and G is matrix-convex in the sense of Bonnans and Shapiro [7, Section 5.3.2], a KKT point is a global optimum of (1.1).

There are several equivalent reformulations for the semidefinite complementarity condition (2.2), among which the simplest one is

$$G(x)Y = O,\ G(x) \in \mathbb{S}^m_+,\ Y \in \mathbb{S}^m_+, \tag{2.5}$$

and two other formulations are

$$\mathrm{Sym}\,(G(x)Y) = O,\ G(x) \in \mathbb{S}^m_+,\ Y \in \mathbb{S}^m_+, \tag{2.6}$$

$$G(x)^{\frac{1}{2}} Y G(x)^{\frac{1}{2}} = O,\ G(x) \in \mathbb{S}^m_+,\ Y \in \mathbb{S}^m_+. \tag{2.7}$$

Based on the above two formulations, primal-dual interior point methods (PDIPMs) have been developed for solving NSDPs so far. For example, see Yamashita et al. [64] and Yamashita and Yabe [63] for PDIPM with (2.6) and also see Okuno [41] for the one with (2.7).

Other fundamental properties of the complementarity condition Let $x^* \in \mathbb{R}^n$ be a KKT point for the NSDP. With an appropriate orthogonal matrix $P_* \in \mathbb{R}^{m\times m}$, the matrix $G(x^*)$ and an arbitrary Lagrange multiplier matrix $Y_* \in \mathcal{Y}(x^*)$ are factorized as

$$G(x^*) = P_* \begin{bmatrix} O & O \\ O & G_*^{\mathrm{FF}} \end{bmatrix} P_*^\top, \ Y_* = P_* \begin{bmatrix} Y_*^{\mathrm{EE}} & O \\ O & O \end{bmatrix} P_*^\top, \tag{2.8}$$

where $G_*^{\mathrm{FF}} \in S_{++}^{r_*}$ is a *diagonal* matrix with $r_* := \mathrm{rank} G(x^*)$ such that the positive real eigenvalues of $G(x^*)$ are aligned on the diagonal line, and $Y_*^{\mathrm{EE}} \in S_+^{m-r_*}$. Without loss of generality, we may assume that the eigenvalues are placed in the ascending order on the diagonal. Needless to say, P_* is a matrix whose columns are eigenvectors of $G(x^*)$. Partition the matrix P_* as

$$P_* = [E_*, F_*],$$

where $E_* \in \mathbb{R}^{m\times(m-r_*)}$ and $F_* \in \mathbb{R}^{m\times r_*}$. Note that each column of E_* represents an eigenvector of $G_* := G(x^*)$ which corresponds to the zero-eigenvalue of $G(x^*)$, as well as that of F_* does to a positive eigenvalue of G_*. In terms of E_* and F_*, the two equations in (2.8) are transformed as

$$\begin{bmatrix} E_*^\top G_* E_* & E_*^\top G_* F_* \\ F_*^\top G_* E_* & F_*^\top G_* F_* \end{bmatrix} = \begin{bmatrix} O & O \\ O & G_*^{\mathrm{FF}} \end{bmatrix}, \ \begin{bmatrix} E_*^\top Y_* E_* & E_*^\top Y_* F_* \\ F_*^\top Y_* E_* & F_*^\top Y_* F_* \end{bmatrix} = \begin{bmatrix} Y_*^{\mathrm{EE}} & O \\ O & O \end{bmatrix}. \tag{2.9}$$

We will often make use of formulation (2.9). For later use, we define the following notations: For the above $P_* = [E_*, F_*]$ and given $x, d \in \mathbb{R}^n$ and $Y \in \mathbb{S}^m$, we write

$$\begin{bmatrix} Y^{\mathrm{EE}} & Y^{\mathrm{EF}} \\ Y^{\mathrm{FE}} & Y^{\mathrm{FF}} \end{bmatrix} := \begin{bmatrix} E_*^\top Y E_* & E_*^\top Y F_* \\ F_*^\top Y E_* & F_*^\top Y F_* \end{bmatrix}, \ \begin{bmatrix} G^{\mathrm{EE}} & G^{\mathrm{EF}} \\ G^{\mathrm{FE}} & G^{\mathrm{FF}} \end{bmatrix} := \begin{bmatrix} E_*^\top G(x) E_* & E_*^\top G(x) F_* \\ F_*^\top G(x) E_* & F_*^\top G(x) F_* \end{bmatrix}, \tag{2.10}$$

$$\begin{bmatrix} \Delta G^{\mathrm{EE}}(x;d) & \Delta G^{\mathrm{EF}}(x;d) \\ \Delta G^{\mathrm{FE}}(x;d) & \Delta G^{\mathrm{FF}}(x;d) \end{bmatrix} := \begin{bmatrix} E_*^\top \Delta G(x;d) E_* & E_*^\top \Delta G(x;d) F_* \\ F_*^\top \Delta G(x;d) E_* & F_*^\top \Delta G(x;d) F_* \end{bmatrix}. \tag{2.11}$$

2.2. Second-order optimality conditions and relevant properties

In this subsection, we review the second-order necessary/sufficient conditions for the NSDP. Subsequently, we will describe the relevant properties briefly. For more detailed explanations, we refer readers to, e.g., [63, 47] or [7].

Definition 3. Let x^* be a KKT point for the NSDP and consider the corresponding Lagrange multiplier set $\Lambda(x^*)$. Then, the nondegeneracy condition, strict complementarity conditions, and second-order condition are defined as follows:

Nondegeneracy condition: Let $r_* := \mathrm{rank}\, G(x^*)$ and let $\{e_1, e_2, \ldots, e_{m-r_*}\}$ be an orthonormal basis of the null space of $G(x^*)$. Moreover, denote

$$v_{ij} := (e_i^\top \mathcal{G}_1(x^*) e_j, \cdots, e_i^\top \mathcal{G}_n(x^*) e_j)^\top \in \mathbb{R}^n \ (1 \le i \le j \le m - r_*).$$

We say that the nondegeneracy condition holds at x^* if the vectors $v_{ij} \in \mathbb{R}^n$ $(1 \le i \le j \le m - r_*)$ and $\nabla h_i(x^*)$ $(i = 1, 2, \ldots, \ell)$ are linearly independent.

Strict complementarity condition: Let $Y_* \in \mathbb{S}_+^m$ be a Lagrange multiplier matrix at x^*, implying that $G(x^*)$ and Y_* satisfies the complementarity condition (2.2). We say that the strict complementarity condition holds at (x^*, Y_*) if $G(x^*) + Y_* \in \mathbb{S}_{++}^m$, which is equivalent to $\mathrm{rank}\, G(x^*) + \mathrm{rank}\, Y_* = m$ under (2.2).

Second-order conditions: We say that the second-order necessary (resp., sufficient) condition holds at x^* if

$$\sup_{(Y,z)\in\Lambda(x^*)} d^\top \left(\nabla^2_{xx}L(x^*,Y,z)+\Omega(x^*,Y)\right)d \geq (\text{resp.},\ >)0 \ \ \forall d\in C(x^*), \tag{2.12}$$

where $C(x^*)$ is the critical cone at x^* and specifically represented as

$$C(x^*)=\left\{d\in\mathbb{R}^n \mid \nabla f(x^*)^\top d=0, \nabla h(x^*)^\top d=0,\ \Delta G(x^*;d)\in T_{\mathbb{S}^m_+}(G(x^*))\right\}. \tag{2.13}$$

Here, $T_{\mathbb{S}^m_+}(G(x^*))$ denotes the tangent cone of $\mathbb{S}^m_+$ at $G(x^*)$ and is represented specifically as

$$T_{\mathbb{S}^m_+}(G(x^*))=\left\{X\in\mathbb{S}^m \mid E_*^\top X E_*(=X^{\mathrm{EE}})\in S^{r_*}_+\right\}. \tag{2.14}$$

Moreover, for any $x\in\mathbb{R}^n$ and $Y\in\mathbb{S}^m$, $\Omega(x,Y)$ denotes the matrix in S^n whose (i,j)-th entry is given as

$$(\Omega(x,Y))_{i,j} := 2Y\bullet \mathcal{G}_i(x)G(x)^\dagger \mathcal{G}_j(x)$$

for $i,j=1,2,\ldots,n$, where $G(x)^\dagger$ denotes the Moore-Penrose inverse matrix of $G(x)$.

Remark 2. The nondegeneracy condition at x^* is a constraint qualification for the NSDP and yields the MFCQ. It reduces to the linear independence constraint qualification (LICQ) when nonlinear optimization is considered. As with the LICQ, the Lagrange multiplier set $\Lambda(x^*)$ is a singleton under the nondegeneracy condition.

The term $d^\top\Omega(x^*,Y)d$ in (3.1) is called the *sigma term* for the semi-definite constraint $G(x)\in\mathbb{S}^m_+$. We refer readers to [7] for the precise description of its background and properties. In the following lemma, the sigma term is expressed more specifically, thereby being ensured to be nonnegative.

Lemma 1. *For a KKT triplet $w^*=(x^*,Y_*,z^*)$ and a direction $d\in\mathbb{R}^n$, it holds that*

$$\begin{aligned} d^\top\Omega(x^*,Y_*)d &= 2\mathrm{Tr}\left(Y_*^{\mathrm{EE}}\Delta G^{\mathrm{FE}}(x^*;d)(G_*^{\mathrm{FF}})^{-1}\Delta G^{\mathrm{EF}}(x^*;d)\right)\\ &= 2\left\|(Y_*^{\mathrm{EE}})^{\frac{1}{2}}\Delta G^{\mathrm{FE}}(x^*;d)(G_*^{\mathrm{FF}})^{-\frac{1}{2}}\right\|_{\mathrm{F}}^2, \end{aligned}$$

where Y_*^{EE} and G_*^{FF} are defined in (2.9), and moreover ΔG^{FE} and ΔG^{EF} in (2.11).

Proof. It is done by straightforward calculation. See Appendix A.1 for details. □

When we consider the standard nonlinear optimization where the nonnegative cone is set in the NSDP in place of the semidefinite cone, the sigma term always vanishes because $\Delta G^{\mathrm{FE}}(x^*;d)=O$ holds for any d in the above lemma, and thus it never appears in the second-order conditions. In contrast, in the NSDP, the sigma term reflects curvature of $\mathbb{S}^m_+$ and is nonnegative for any $d\neq 0$ and $Y\in\mathbb{S}^m_+$ as shown in Lemma 1. With the help of this term, the second-order condition is more likely to hold even when $\nabla^2_{xx}L$ is not positive semidefinite over the critical cone. Meanwhile, its complicated structure often brings about difficulty of analyzing several properties which were already shown only for nonlinear optimization.

Lastly, we mention useful facts associated with the second-order conditions in the following two necessary and sufficient optimality conditions.

Second-order necessary optimality for the NSDP Let $x^*\in\mathbb{R}^n$ be a local optimum of NSDP (1.1) and suppose that the MFCQ holds there. Then, the second-order condition holds at x^*.

Second-order sufficient optimality for the NSDP Suppose that x^* is a KKT point of NSDP (1.1) and, furthermore, the second-order sufficient condition holds. Then, x^* is a strict local optimum of NSDP (1.1). In particular, the quadratic growth condition holds, that is, there exists some $q>0$ and vicinity $\mathcal{N}(x^*)$ of x^* such that $f(x)-f(x^*)\geq q\|x-x^*\|^2$ for all $x\in\mathcal{N}(x^*)\cap\mathcal{F}$.

2.3. BKKT conditions and central path

In this section, we formally define the central path for the NSDP by introducing the barrier KKT (BKKT) conditions. The BKKT conditions are composed of (2.1), (2.3), and the following perturbed conditions for (2.5): For $\mu > 0$,

$$G(x)Y = \mu I,\ G(x) \in \mathbb{S}^m_{++},\ Y \in \mathbb{S}^m_{++}. \tag{2.15}$$

The parameter μ is often refereed to as barrier parameter, and x and (x, Y, z) satisfying the BKKT conditions are called a BKKT point and BKKT triplet, respectively. It is worth mentioning that condition (2.15) is equivalent to the one obtained by replacing O with μI in (2.6) or (2.7). As μ gets closer to 0, BKKT points are expected to approach the set of KKT points for the NSDP. A basic algorithmic policy of primal-dual interior point methods is to track BKKT triplets together with driving μ to 0, so as to reach a KKT triplet.

We refer to a path formed by BKKT triplets as a *central path*.

3. Main analysis

3.1. Assumptions and outline of analysis

Throughout Section 3, x^* denotes a KKT point of the NSDP, and is assumed to satisfy the following:

Assumption 1. *The KKT point x^* satisfies the following three conditions:*
1. *There exists a Lagrange multiplier matrix $Y_* \in \mathbb{S}^m_+$ satisfying the strict complementarity condition.*
2. *The strong second-order sufficient condition (SSOSC) holds: For all $(Y, z) \in \Lambda(x^*)$, it holds that*

$$d^\top \left(\nabla^2_{xx} L(x^*, Y, z) + \Omega(x^*, Y)\right) d > 0 \ \ \forall d \in C(x^*) \setminus \{0\}. \tag{3.1}$$

3. *The MFCQ holds at x^*.*

Under the above SSOSC, x^* is a strict local optimum of the NSDP. The SSOSC holds, for example, when f is strongly convex and G and h are affine. It is worth mentioning that this SSOSC differs from the one proposed by Sun [49]. It is regarded as a straightforward generalization of the SSOSC considered by Wright and Orban [58] for nonlinear optimization.

The MFCQ ensures compactness and convexity of the Lagrange multiplier set $\Lambda(x^*)$, but $\Lambda(x^*)$ is not necessarily a singleton. This is a quite different situation from the case where the nondegeneracy condition is supposed. (cf. Remark 2)

Goal and outline of the analysis: The goal of the whole analysis we will conduct is to prove that under the above assumptions, there exists a unique and smooth central path converging to the KKT triplet

$$w^{\mathrm{a}} := (x^*, Y_{\mathrm{a}}, z^{\mathrm{a}}), \tag{3.2}$$

where $(Y_{\mathrm{a}}, z^{\mathrm{a}}) \in \Lambda(x^*)$ is called an analytic center at x^*, defined formally in the next subsection. In order to achieve this goal, we will prove the following claims in order:

Claim (i): There exists a sequence of BKKT triplets $\{w^k = (x^k, Y_k, z^k)\}$ converging to the KKT triplet w^{a} (cf. Theorem 1 in subsection 3.3).

Claim (ii): A direction of BKKT point x^k approaching the KKT point x^* converges to some direction ξ^*. This ξ^* is a unique x-component solution of a certain linear equation related to tangential directions of the central path (cf. Theorem 2 and Corollary 1 in subsection 3.4).

Claim (iii): For any barrier parameter μ small enough, a corresponding BKKT point exists uniquely in a certain convex region that admits x^* as a vertex and ξ^* as an axis (cf. Theorem 3 in subsection 3.6).

With the help of the above claims and the classical implicit function theorem, we will prove our main claim of the goal (cf. Theorem 4 in subsection 3.7). Mind that henceforth, several proofs are deferred to the Appendix for the sake of readability.

Additional notations and symbols used hereafter: In the remaining of Section 3 and the Appendix, we will use the symbols and the notations defined in (2.8)-(2.11) in addition to those introduced at the end of Section 1. In particular, $P_* = [E_*, F_*]$ is an *arbitrarily* chosen orthogonal matrix defined for $G(x^*)$ so that (2.8) holds. For the sake of simplicity, we often write

$$G_* := G(x^*),\ G_k := G(x^k).$$

Besides, we will make use of G_*^{ind} and G_k^{ind} ($ind \in \{\mathrm{EE}, \mathrm{FE}, \mathrm{EF}, \mathrm{FF}\}$) defined by replacing G and Y in (2.10) with $G_*$and G_k, respectively. Furthermore, Y_k^{ind} and Y_{a}^{ind} ($ind \in \{\mathrm{EE}, \mathrm{FE}, \mathrm{EF}, \mathrm{FF}\}$) are defined in a similar way using Y_k and Y_{a}.

3.2. Existence of analytic center for NSDP

The analytic center for the NSDP at x^* is formally defined as follows:

Definition 4 (Analytic center for NSDP (1.1)). We say that $(Y_{\mathrm{a}}, z^{\mathrm{a}}) \in \Lambda(x^*)$ is an analytic center of NSDP (1.1) at x^* if it is an optimum of

$$\min\ -\log\det Y^{\mathrm{EE}} \ \text{s.t.}\ (Y, z) \in \Lambda(x^*). \tag{3.3}$$

Here, we define $\log 0 := -\infty$ by convention.

In the next proposition, we ensure existence and uniqueness of the analytic center at x^*. In other words, the KKT triplet w^{a} defined in (3.2) is well-defined.

Proposition 1. *Suppose that Assumption 1 holds. Then, an analytic center of NSDP (1.1) at x^* exists uniquely. In particular, $(Y_{\mathrm{a}}, z^{\mathrm{a}}) \in \mathbb{S}^m \times \mathbb{R}^s$ is the analytic center at x^* if and only if $(Y_{\mathrm{a}}, z^{\mathrm{a}}) \in \Lambda(x^*)$ and there exists some vector $v \in \mathbb{R}^n$ such that*

$$\Delta G^{\mathrm{EE}}(x^*; v) = (Y_{\mathrm{a}}^{\mathrm{EE}})^{-1},\ \nabla h(x^*)^\top v = 0. \tag{3.4}$$

Proof. See Appendix A.2. □

3.3. Proof of Claim (i): Convergence of BKKT triplets to KKT triplet with analytic center

In this subsection, we will prove that there exists a sequence of BKKT points which converges to the KKT point x^*. Moreover, we will show that the corresponding dual sequence converges to the analytic center $(Y_{\mathrm{a}}, z^{\mathrm{a}})$.

Let us define the following log-barrier function for the NSDP: for each $\mu > 0$

$$\psi_\mu(x) := f(x) - \mu \log\det G(x).$$

The following proposition states that there exists a sequence of local optima of barrier penalized NSDPs converging to x^*. Such local optima are BKKT points of the NSDP locally around x^*.

Proposition 2. *Let Assumption 1 hold and $\{\mu_k\} \subseteq \mathbb{R}_{++}$ be an arbitrary decreasing sequence converging to 0. Then, there exists a sequence $\{x^k\} \subseteq \mathbb{R}^n$ such that $\lim_{k\to\infty} x^k = x^*$ and, for any $k \geq \overline{K}$ with $\overline{K}$ sufficiently large, x^k is a local optimum of*

$$\min\ \psi_{\mu_k}(x)\ \text{ s.t. }\ h(x) = 0,\ G(x) \in \mathbb{S}^m_{++}. \tag{3.5}$$

Proof. The proof is analogous to those of classical results as to penalty methods [35], although it is different in dealing with the log determinant function and the semidefinite constraint. Nonetheless, the precise proof is given in Appendix A for completeness. □

In Proposition 2, as $\nabla h(x^*)$ is of full column rank, so is $\nabla h(x^k)$ for any $k \geq \overline{K}$ with $\overline{K}$ large enough, and thus the KKT conditions for (3.5) holds at x^k. From the KKT conditions together with $\nabla \psi_\mu(x) = \nabla f(x) - \mu \mathcal{J} G(x)^* G(x)^{-1}$, there exists $z^k \in \mathbb{R}^s$ such that

$$\nabla f(x^k) - \mu_k \mathcal{J} G(x^k)^* G_k^{-1} + \nabla h(x^k) z^k = 0, \ h(x^k) = 0, \ G_k \in \mathbb{S}^m_{++},$$

which together with $Y_k := \mu_k G_k^{-1} \in \mathbb{S}^m_{++}$ implies that x^k and $w^k := (x^k, Y_k, z^k) \in \mathcal{W}_{++}$ are BKKT point and BKKT triplet for each $k \geq \overline{K}$, respectively.

In summary, as a consequence of Proposition 2, given a decreasing sequence $\{\mu_k\} \subseteq \mathbb{R}_{++}$ converging to 0, there exists an integer $\overline{K} > 0$ and $\{w^k\} \subseteq \mathcal{W}_{++}$ with $w^k = (x^k, Y_k, z^k)$ such that

$$\lim_{k\to\infty} x^k = x^*, \ Y_k = \mu_k G_k^{-1} \in \mathbb{S}^m_{++}$$

and, for each $k \geq \overline{K}$,

$$\nabla h(x^k): \text{ full column rank}, \ w^k: \text{ BKKT triplet with barrier parameter } \mu_k.$$

In what follows, for the sake of brevity, we assume $\overline{K} = 0$. Moreover, we suppose that $x^k \neq x^*$ for all k without loss of generality and define

$$d^k := x^k - x^*. \tag{3.6}$$

Notice that $\lim_{k\to\infty} d^k = 0$ by construction. Hereafter, we focus on those sequences $\{w^k\}$ and $\{d^k\}$.

Remark 3. In a quite similar manner to the proof of Yamashita et al. [64, Theorem 1], we ensure that, under the MFCQ at x^*, the sequence $\{(Y_k, z^k)\}$ is bounded, and its accumulation point together with x^* fulfills the KKT conditions of NSDP (1.1).

In fact, the whole sequence $\{(Y_k, z^k)\}$ converges to the analytic center $(Y_{\mathrm{a}}, z^{\mathrm{a}})$. To prove this claim, we first present the following proposition implying that the convergence speeds of $\{\mu_k\}$ and $\{\|d^k\|\}$ are equivalent.

Proposition 3. *Suppose that Assumption 1 holds. Then, we have*

$$\mu_k = \Theta(\|d^k\|).$$

Proof. See Appendix A.4. □

Using this proposition, the convergence to the analytic center can be established.

Theorem 1. *Suppose that Assumption 1 holds. Then, the whole sequence* $\{(Y_k, z^k)\}$ *converges to the analytic center* $(Y_{\mathrm{a}}, z^{\mathrm{a}})$ *of NSDP (1.1) at* x^*.

Proof. For each $k \geq 0$, let $\widetilde{d^k} := \frac{d^k}{\|d^k\|}$. Since $\{\widetilde{d^k}\}$ is bounded, it has at least one accumulation point, say $\widetilde{d^*}$. Choose an arbitrary subsequence $\{\widetilde{d^k}\}_{k\in\mathcal{K}}$ which converges to $\widetilde{d^*}$. From Proposition 3, $\left\{\frac{\mu_k}{\|d^k\|}\right\}_{k\in\mathcal{K}}$ is bounded and any accumulation point, say $\bar{\alpha} \in \mathbb{R}$, is positive. Without loss of generality, we assume that $\left\{\frac{\mu_k}{\|d^k\|}\right\}_{k\in\mathcal{K}}$ converges to $\bar{\alpha} > 0$ by taking a subsequence further if necessary.

Note that $\|Y_k\|_{\mathrm{F}} = \mathrm{O}(1)$ as was explained in Remark 3. Recalling that $P_* = \begin{bmatrix} E_*, & F_* \end{bmatrix}$ is orthogonal, we have

$$\begin{aligned} \frac{\mu_k I_m}{\|d^k\|} &= \frac{P_*^\top G_k Y_k P_*}{\|d^k\|} \\ &= \frac{P_*^\top \left(G(x^*) + \Delta G(x^*; d^k) + \mathrm{O}(\|d^k\|^2)\right) P_* P_*^\top Y_k P_*}{\|d^k\|} \\ &= \begin{bmatrix} O & O \\ \frac{1}{\|d^k\|} G_*^{\mathrm{FF}} Y_k^{\mathrm{EF}} & \frac{1}{\|d^k\|} G_*^{\mathrm{FF}} Y_k^{\mathrm{FF}} \end{bmatrix} + P_*^\top \Delta G(x^*; \widetilde{d^k}) P_* \left(P_*^\top Y_k P_*\right) + \mathrm{O}(\|d^k\|). \end{aligned} \tag{3.7}$$

Taking into account that $\lim_{k\in\mathcal{K}\to\infty} P_*^\top Y_k P_* = \begin{bmatrix} Y_*^{\mathrm{EE}} & O \\ O & O \end{bmatrix}$ with $Y_*^{\mathrm{EE}} = E_*^\top Y_* E_*$ and driving $k \in \mathcal{K} \to \infty$ in the $(1,1)$-block component of (3.7), we obtain $\Delta G^{\mathrm{EE}}(x^*; \widetilde{d^*}) Y_*^{\mathrm{EE}} = \bar{\alpha} I_{r_*}$, implying

$$(Y_*^{\mathrm{EE}})^{-1} = \Delta G^{\mathrm{EE}}(x^*; \bar{\alpha}^{-1}\widetilde{d^*}). \tag{3.8}$$

Moreover, for each $k \in \mathcal{K}$, it holds that

$$0 = \frac{h(x^k)}{\|d^k\|} = \frac{h(x^*) + \nabla h(x^*)^\top d^k + \mathrm{O}(\|d^k\|^2)}{\|d^k\|} = \nabla h(x^*)^\top \widetilde{d^k} + \mathrm{O}(\|d^k\|),$$

which along with driving $k (\in \mathcal{K}) \to \infty$ and multiplying $\bar{\alpha}^{-1}$ implies $\nabla h(x^*)^\top (\bar{\alpha}^{-1}\widetilde{d^*}) = 0$. Comparing this fact and (3.8) to condition (3.4) with $v := \bar{\alpha}\widetilde{d^*}$, we ensure that (Y_*, z^*) is an analytic center of the NSDP at x^*, leading to $(Y_{\mathrm{a}}, z^{\mathrm{a}}) = (Y_*, z^*)$ due to the uniqueness of analytic center by Proposition 1. Finally, recalling that (Y_*, z^*) is an arbitrary accumulation point of $\{(Y_k, z^k)\}$, we conclude that the whole sequence $\{(Y_k, z^k)\}$ converges to $(Y_{\mathrm{a}}, z^{\mathrm{a}})$. The proof is complete. □

Before moving on to the next subsection, we show that $\|Y_k^{\mathrm{EF}}\|_{\mathrm{F}}$ and $\|Y_k^{\mathrm{FF}}\|_{\mathrm{F}}$ are bounded by $\mathrm{O}(\mu_k)$.

Proposition 4. *Suppose that Assumption 1 holds. Then, we have*

$$\|Y_k^{\mathrm{EF}}\|_{\mathrm{F}} = \mathrm{O}(\mu_k),\ \|Y_k^{\mathrm{FF}}\|_{\mathrm{F}} = \mathrm{O}(\mu_k).$$

Proof. Note that $\{(Y_k, z^k)\}$ is convergent by Theorem 1. Applying Taylor's expansion to G_k around x^* and using $G_k Y_k = \mu_k I$ and $G_* Y_* = G_* Y_{\mathrm{a}} = O$ give

$$\begin{aligned} \mu_k I &= \left(G_* + \Delta G(x^*; d^k) + \mathrm{O}(\|d^k\|^2)\right)(Y_k - Y_{\mathrm{a}} + Y_{\mathrm{a}}) \\ &= G_*(Y_k - Y_{\mathrm{a}}) + \mathrm{O}(\|d^k\|) \\ &= P_*^\top \begin{bmatrix} O & O \\ G_*^{\mathrm{FF}} Y_k^{\mathrm{FE}} & G_*^{\mathrm{FF}} Y_k^{\mathrm{FF}} \end{bmatrix} P_* + \mathrm{O}(\|d^k\|). \end{aligned}$$

Recall that P_* is an orthogonal matrix. Divide both the sides of the above by μ_k and drive $k \to \infty$. From Proposition 3, we obtain

$$\frac{\|G_*^{\mathrm{FF}} Y_k^{\mathrm{FE}}\|_{\mathrm{F}}}{\mu_k} = \mathrm{O}(1),\ \frac{\|G_*^{\mathrm{FF}} Y_k^{\mathrm{FF}}\|_{\mathrm{F}}}{\mu_k} = \mathrm{O}(1),$$

which together with $G_*^{\mathrm{FF}} \in \mathbb{S}_{++}^{m-r_*}$ implies the desired assertions. □

3.4. Proof of Claim (ii): Asymptotic behavior of directions of BKKT points approaching KKT point

Let $\{w^k = (x^k, Y_k, z^k)\} \subseteq \mathcal{W}_{++}$ be a sequence of BKKT triplets described right after Proposition 2. From Theorem 1, $\{w^k\}$ converges to the KKT triplet $w^* = (x^*, Y_{\mathrm{a}}, z^{\mathrm{a}})$ with analytic center $(Y_{\mathrm{a}}, z^{\mathrm{a}})$. In this subsection, we study how d^k/μ_k behaves asymptotically, wherein d^k is defined in (3.6).

We begin by considering the following equation-system that comes from the BKKT conditions of the symmetric form:

$$\nabla_x L(w) = 0,\ G(x)Y + YG(x) = 2\mu I,\ h(x) = 0, \tag{3.9}$$

where $w = (x, Y, z) \in \mathcal{W}_{++}$. Suppose at this moment[1] that there exists a smooth function $w(\cdot): (0, \bar{\mu}] \to \mathcal{W}_{++}$ with some $\bar{\mu} > 0$ such that $w(\mu)$ is a BKKT triplet for each $\mu \in (0, \bar{\mu}]$ and we stand at $w = w(\mu)$. Differentiating equations (3.9) with respect to μ results in

$$\nabla_{xx}^2 L(w)\dot{x} - \mathcal{J}G(x)^* \dot{Y} + \nabla h(x)\dot{z} = 0, \tag{3.10}$$

$$\mathcal{L}_{G(x)}\dot{Y} + \mathcal{L}_Y \Delta G(x; \dot{x}) = 2I, \tag{3.11}$$

$$\nabla h(x)^\top \dot{x} = 0. \tag{3.12}$$

[1] In Theorem 4, this assumption will be verified.

As for the definition of $\mathcal{L}_{(\cdot)}(\cdot)$, refer to the section of notations. For later use, in terms of the matrix function

$$\mathcal{A}(w) := \begin{bmatrix} \nabla^2_{xx} L(w) & -\mathcal{J}G(x)^* & \nabla h(x) \\ \mathcal{L}_Y \mathcal{G}_1(x) \cdots \mathcal{L}_Y \mathcal{G}_n(x) & \mathcal{L}_{G(x)} & 0 \\ \nabla h(x)^\top & 0 & 0 \end{bmatrix}, \tag{3.13}$$

we express the above equation-system (3.10)-(3.12) as

$$\mathcal{A}(w)\dot{w} = \begin{bmatrix} 0 \\ 2I \\ 0 \end{bmatrix}.$$

Now, relevant to this equation-system, we consider the following equations defined at the KKT triplet $w^{\mathrm{a}} = (x^*, Y_{\mathrm{a}}, z^{\mathrm{a}})$:

$$U_{x^*}^\top (\nabla^2_{xx} L(w^{\mathrm{a}})\Delta x - \mathcal{J}G(x^*)^* \Delta Y) = 0, \tag{3.14}$$
$$\mathcal{L}_{G(x^*)}\Delta Y + \mathcal{L}_{Y_{\mathrm{a}}} \Delta G(x^*; \Delta x) = 2I, \tag{3.15}$$
$$\nabla h(x^*)^\top \Delta x = 0, \tag{3.16}$$

where U_{x^*} denotes an arbitrary matrix whose columns form a normal orthogonal basis of the subspace

$$\mathcal{U}_* := \{d \in \mathbb{R}^n \mid \Delta G^{\mathrm{EE}}(x^*; d) = O, \nabla h(x^*)^\top d = 0\} \tag{3.17}$$

and we can write $U_{x^*} \in \mathbb{R}^{n \times p_*}$ by letting p_* be the dimension of $\mathcal{U}_*$. Notice that the above equations (3.14)-(3.16) are derived by changing the variables in (3.10)-(3.12), pre-multiplying (3.10) by the matrix $U_{x^*}^\top$, and using the relation $\nabla h(x^*)^\top U_{x^*} = 0$. The following proposition holds as to the solution set of equations (3.14)-(3.16):

Proposition 5. *Suppose that Assumption 1 holds. Let*

$$S := \{(\Delta x, \Delta Y) \in \mathbb{R}^n \times \mathbb{S}^m \text{: solution to (3.14)-(3.16)}\}.$$

If $S \neq \emptyset$, then the following properties hold:
1. *Δx-component in S is unique, written as $\xi^* \in \mathbb{R}^n$;*
2. *$\Delta Y^{\mathrm{FF}} = (G_*^{\mathrm{FF}})^{-1}$, $\Delta G^{\mathrm{EE}}(x^*; \xi^*) = (Y_{\mathrm{a}}^{\mathrm{EE}})^{-1}$, $\Delta Y^{\mathrm{EF}} = -Y_{\mathrm{a}}^{\mathrm{EE}} \Delta G^{\mathrm{EF}}(x^*; \xi^*)(G_*^{\mathrm{FF}})^{-1}$.*

Proof. See Appendix A.5. □

The following theorem shows that the limit of d^k/μ_k is actually equal to the direction ξ^*, which is defined in the above proposition.

Theorem 2. *Suppose that Assumption 1 holds. Let d^k be the vector defined in (3.6) and ξ^* be the one defined in Proposition 5. Then, we have*

$$\lim_{k \to \infty} \frac{d^k}{\mu_k} = \xi^*.$$

In particular, $\xi^* \neq 0$.

Proof. From Proposition 3, $\{d^k/\mu_k\}$ is bounded. Let $\tilde{\xi} \in \mathbb{R}^n$ be an arbitrary accumulation point of $\{d^k/\mu_k\}$. In order to prove the assertion, it suffices to show that $\tilde{\xi}$ is a Δx-component of the solution set of equations (A.27)-(A.29) because of item 1 of Proposition 5. Altering Y_k as

$$\widehat{Y}_k := P_* \begin{bmatrix} Y_{\mathrm{a}}^{\mathrm{EE}} & Y_k^{\mathrm{EF}} \\ Y_k^{\mathrm{FE}} & Y_k^{\mathrm{FF}} \end{bmatrix} P_*^\top$$

for each k, we obtain

$$\begin{aligned}
\|\widehat{Y}_k - Y_{\rm a}\|_{\rm F} &= \left\| \begin{bmatrix} Y_{\rm a}^{\rm EE} - Y_{\rm a}^{\rm EE} & Y_k^{\rm EF} - Y_{\rm a}^{\rm EF} \\ Y_k^{\rm FE} - Y_{\rm a}^{\rm FE} & Y_k^{\rm FF} - Y_{\rm a}^{\rm FF} \end{bmatrix} \right\|_{\rm F} \\
&= \left\| \begin{bmatrix} O & Y_k^{\rm EF} \\ Y_k^{\rm FE} & Y_k^{\rm FF} \end{bmatrix} \right\|_{\rm F} \\
&= {\rm O}(\|Y_k^{\rm EF}\|_{\rm F} + \|Y_k^{\rm FF}\|_{\rm F}) \\
&= {\rm O}(\mu_k),
\end{aligned}$$

where the last equality follows from Proposition 4, and thus $\left\{\frac{1}{\mu_k}(\widehat{Y}_k - Y_{\rm a})\right\}$ is bounded and has at least one accumulation point, say ΔY_*. By the BKKT conditions and applying Taylor's expansion to $\nabla_x L(w^k)$ around w^*, it holds that

$$\begin{aligned}
0 &= \frac{1}{\mu_k} U_{x^*}^\top \nabla_x L(w^k) \\
&= \frac{1}{\mu_k} U_{x^*}^\top \left(\nabla_{xx}^2 L(w^{\rm a}) d^k - \mathcal{J}G(x^*)^*(Y_k - Y_{\rm a})\right) + {\rm O}(\|d^k\|^2) \\
&= U_{x^*}^\top \left(\nabla_{xx}^2 L(w^{\rm a}) \frac{d^k}{\mu_k} - \mathcal{J}G(x^*)^* \frac{(\widehat{Y}_k - Y_{\rm a})}{\mu_k}\right) + \frac{1}{\mu_k}{\rm O}(\|d^k\|^2),
\end{aligned}$$

where the first equality follows from $\nabla_x L(w^k) = 0$, the second one does from $\nabla_x L(w^{\rm a}) = 0$ and $U_{x^*}^\top \nabla h(x^*) = 0$, and the third one does from $U_{x^*}^\top \mathcal{J}G(x^*)^* Y_k = U_{x^*}^\top \mathcal{J}G(x^*)^* \widehat{Y}_k$. Driving $k \to \infty$ above and using Proposition 3 imply

$$U_{x^*}^\top \left(\nabla_{xx}^2 L(w^{\rm a}) \tilde{\xi} - \mathcal{J}G(x^*)^* \Delta Y^*\right) = 0,$$

which is nothing but (3.14) with $(\Delta x, \Delta Y) = (\widetilde{\xi}, \Delta Y_*)$.

Next, by $\mathcal{L}_{G_k} Y_k + \mathcal{L}_{Y_k} G_k = 2\mu_k I$ from the BKKT conditions and also by noting $G_* \widehat{Y}_k = G_* Y_k$ together with $G_* Y_{\rm a} = O$, there holds that

$$\begin{aligned}
I &= \frac{1}{\mu_k}(G_k Y_k - G_* Y_{\rm a}) \\
&= \frac{1}{\mu_k}\left(\left(G_* + \Delta G(x^*; d^k) + {\rm O}(\|d^k\|^2)\right) Y_k - G_* Y_{\rm a}\right) \\
&= \frac{1}{\mu_k}\left(G_*(\widehat{Y}_k - Y_{\rm a}) + \Delta G(x^*; d^k) Y_k\right) + {\rm O}(\|d^k\|),
\end{aligned}$$

wherein by driving $k \to \infty$, symmetrizing, and using $\lim_{k\to\infty} Y_k = Y_{\rm a}$, we gain (3.15) with $(\Delta x, \Delta Y) = (\widetilde{\xi}, \Delta Y_*)$. Finally, we can prove (3.12) with $(\Delta x, \Delta Y) = (\widetilde{\xi}, \Delta Y_*)$ by driving k to ∞ in the relation $0 = \frac{1}{\mu_k}(h(x^k) - h(x^*)) = \frac{1}{\mu_k}\nabla h(x^*)^\top d^k + {\rm O}(\|d^k\|)$. Consequently, $(\widetilde{\xi}, \Delta Y_*)$ solves (3.12)-(3.15). Hence, $\widetilde{\xi} = \xi^*$ is ensured by using item 1 of Proposition 5. The last assertion follows immediately since $\|d^k\| = \Theta(\mu_k)$ from Proposition 3. The proof is complete. □

It is worth noting that we have multiple choices for $\{(d^k, \mu_k)\}$, while ξ^* is the constant vector that is uniquely determined as a Δx-component of the solution set to the equation-system (3.14)-(3.16). Nevertheless, according to Theorem 2, any $\{d^k/\mu_k\}$ converges to ξ^*.

Theorem 2 yields the following corollary, a clear picture about how x^k approaches x^*.

Corollary 1. *Under Assumption 1, we obtain* $\lim_{k\to\infty} \frac{x^k - x^*}{\|x^k - x^*\|} = \frac{\xi^*}{\|\xi^*\|}$. *This indicates that x^k approaches x^* along the direction $-\xi^*$ asymptotically.*

Proof. From Theorem 2, we have

$$\lim_{k\to\infty} \frac{x^k - x^*}{\|x^k - x^*\|} = \lim_{k\to\infty} \frac{x^k - x^*}{\mu_k} \frac{\mu_k}{\|x^k - x^*\|} = \frac{\xi^*}{\|\xi^*\|}.$$

The proof is complete. □

For $\rho > 0$ and $\mu \geq 0$, define

$$\mathcal{P}_\rho(\mu) := \{x \in \mathbb{R}^n \mid \|x^* + \mu\xi^* - x\| < \rho\mu\|\xi^*\|\}.$$

From Theorem 2, $x^k \in \mathcal{P}_\rho(\mu_k)$ holds for any k sufficiently large. This fact implies that $\{x^k\}_{k\geq K} \subseteq \bigcup_{\mu\geq 0}\mathcal{P}_\rho(\mu)$ for a sufficiently large K. In the next subsection, we study properties of $\mathcal{P}_\rho(\mu)$ more precisely.

3.5. Some properties on $\mathcal{P}_\rho(\mu)$

In this subsection, we will present two propositions about properties on $\mathcal{P}_\rho(\mu)$. In the first proposition, we show a convergence property of $\mu G(x)^{-1}$ for $x \in \mathcal{P}_\rho(\mu)$ when $\mu > 0$ tends to 0.

Proposition 6. *The following properties hold.*

1. *There exist some $\bar{\rho} > 0$ and $\bar{\mu} > 0$ such that $G(x) \in \mathbb{S}^m_{++}$ holds for any $x \in \mathcal{P}_{\bar{\rho}}(\mu)$ and $\mu \leq \bar{\mu}$. Moreover, $\{\mu G(x)^{-1} \in \mathbb{S}^m_{++} \mid x \in \mathcal{P}_{\bar{\rho}}(\mu), \mu \in (0, \bar{\mu}]\}$ is bounded.*
2. *For an arbitrary $\varepsilon > 0$, there exists $\rho > 0$ such that the distance between the matrix Y_{a} and arbitrary accumulation points of $\mu G(x)^{-1}$ $(x \in \mathcal{P}_\rho(\mu))$ is smaller than ε. Particularly when the parameter ρ is varied satisfying $\rho = o(\mu)$, we have*

$$\lim_{\mu\to 0+, \mathcal{P}_\rho(\mu)\ni x\to x^*} \mu G(x)^{-1} = Y_{\mathrm{a}}.$$

Proof. See Appendix A.6. □

Let $\bar{\rho}$ be the constant defined in Proposition 6. Recall that x^k is a BKKT point with barrier parameter μ_k and converges to the KKT point x^* as $k \to \infty$. By the first-half assertion of item of Proposition 6, it holds that

$$\left(\bigcup_{\mu>0}\mathcal{P}_{\bar{\rho}}(\mu) \cap \mathcal{B}_r(x^*)\right) \subseteq \{x \mid G(x) \in \mathbb{S}^m_{++}\}$$

for $r > 0$ sufficiently small, where $\mathcal{B}_r(x^*)$ denotes the closed Euclidean ball centered at x^* with a radius r.

The second proposition will play an important role for proving Theorem 3 that will appear later.

Proposition 7. *Suppose that Assumption 1 holds. Let $\gamma_1 > 0$ and $\gamma_2 > 0$. Choose $\bar{\mu}, \bar{\rho} > 0$ sufficiently small. Then, the following hold:*

1. *For any $w = (x, Y, z) \in \mathbb{R}^n \times \mathbb{S}^m_{++} \times \mathbb{R}^s$ such that*

$$x \in \bigcup_{\mu\in(0,\bar{\mu}]}\mathcal{P}_{\bar{\rho}}(\mu), \tag{3.18}$$
$$\|Y - \mu G(x)^{-1}\|_{\mathrm{F}} \leq \gamma_1\mu, \tag{3.19}$$
$$\|\nabla_x L(w)\| \leq \gamma_2\mu, \tag{3.20}$$

we have

$$d^\top \nabla^2_{xx}L(w)d + \Delta G(x; d) \bullet \mathcal{L}^{-1}_{G(x)}\mathcal{L}_Y(\Delta G(x; d)) > 0, \quad \forall d \in \mathbb{R}^n \setminus \{0\} : \nabla h(x)^\top d = 0. \tag{3.21}$$

See the footnote[2].

2. *The matrix $\mathcal{A}(w)$ defined in (3.13) is nonsingular for any w satisfying conditions (3.18)–(3.20).*

Proof. See Appendix A.7. □

Remark 4. Under $x \in \mathcal{P}_\rho(\mu)$ with sufficiently small $\rho > 0$, conditions (3.19) and (3.20) are induced by $\|G(x)Y - \mu I\|_{\mathrm{F}} = \mathrm{O}(\mu^2)$ and $\|\nabla_x L(w)\| = \mathrm{O}(\mu)$, respectively.

By virtue of item 2 of Proposition 7, an equation-system that admits $\mathcal{A}(w)$ as a coefficient-matrix, such as the one (3.10)-(3.12) and Newton equations relevant to the KKT and BKKT conditions, is ensured to be uniquely solvable as long as w satisfies (3.18)-(3.20). This fact would be useful particularly when considering the Newton method in the primal-dual interior point method.

[2] Note that since $\nabla h(x^*)$ is of full column rank, so is $\nabla h(x)$ for any $x \in \mathcal{P}_\rho(\mu)$ by taking $\mu > 0$ small enough.

3.6. Proof of Claim (iii): Uniqueness of BKKT point for each barrier parameter

Let

$$\mathcal{M} := \{\, x \in \mathbb{R}^n \mid h(x) = 0 \,\}.$$

Under the presence of the full column rank of $\nabla h(x^*)$ and twice continuous differentiability of h, the implicit function theorem yields that there exists an open *ball* $V \subseteq \mathbb{R}^{n-s}$ together with a C^2-*diffeomorphism* $\Phi : V \to \Phi(V) (\subseteq \mathbb{R}^n)$ such that $x^* \in \Phi(V) \subseteq \mathcal{M}$.[3]

Let us give some relevant properties of Φ for later use. Since V is bounded and Φ is smooth, there exists a Lipshidtz constant $M_1 > 0$ such that

$$\|\Phi(u) - \Phi(v)\| \le M_1 \|u - v\|, \quad \forall u, v \in V. \tag{3.22}$$

Moreover, by noting the diffeomorphism of Φ there exists $M_2 > 0$ such that

$$\|\Phi^{-1}(x) - \Phi^{-1}(y)\| \le M_2 \|x - y\|, \quad \forall x, y \in \Phi(V). \tag{3.23}$$

Let

$$\mathrm{dist}\,(x, \mathcal{M}) := \min_{y \in \mathcal{M}} \|x - y\|$$

for $x \in \mathbb{R}^n$. The following lemma holds.

Lemma 2. *It holds that*

$$\mathrm{dist}\,(\check{x}(\mu), \mathcal{M}) = \mathrm{O}(\mu^2), \tag{3.24}$$

where

$$\check{x}(\mu) := x^* + \mu \xi^* \quad (\mu \ge 0).$$

Proof. See Appendix A.8. □

In terms of Φ, NSDP (1.1) is reformulated as the following problem without equality constraints locally around $v_* := \Phi^{-1}(x^*) \in V$:

$$\min_{v \in V} f(\Phi(v)) \ \text{ s.t. } \ G(\Phi(v)) \in \mathbb{S}^m_+.$$

Accordingly, we obtain he following barrier penalized problem for each $\mu > 0$:

$$\min_{v \in V} \Psi_\mu(v) := f(\Phi(v)) - \mu \log\det G(\Phi(v)) \ \text{ s.t. } \ G(\Phi(v)) \in \mathbb{S}^m_{++}. \tag{3.25}$$

Theorem 3. *Take a sufficiently small $\bar{\mu} > 0$. For a barrier parameter $\mu \in (0, \bar{\mu}]$, there exists a unique BKKT point $x(\mu)$ in $\Phi(V) \cap \bigcup_{\mu \in (0,\bar{\mu}]} \mathcal{P}_{\bar{\rho}}(\mu)$. In particular, such $x(\mu)$ is a strict local optimum of problem* (3.25).

Proof. We prove the first assertion by deriving a contradiction. Assume to the contrary that there exists an infinite sequence $\{\mu_\ell\} \subseteq \mathbb{R}_{++}$ which converges to 0 and moreover accompanies two sequences $\{x^\ell\}, \{y^\ell\} \subseteq \mathrm{int}\left(\Phi(V) \cap \bigcup_{\mu \in (0,\mu_\ell]} \mathcal{P}_{\bar{\rho}}(\mu)\right)$ such that for each ℓ, $x^\ell \neq y^\ell$, but x^ℓ and y^ℓ are both BKKT points with barrier parameter μ_ℓ. Hence,

$$v^\ell := \Phi^{-1}(x^\ell), \ \theta^\ell := \Phi^{-1}(y^\ell)$$

exist in V, and moreover

$$\nabla \Psi_{\mu_\ell}(v^\ell) = \nabla \Psi_{\mu_\ell}(\theta^\ell) = 0, \tag{3.26}$$

where Ψ_μ is defined in (3.25). Moreover, we have

$$x^\ell, y^\ell \in \Phi(V) \cap \mathcal{P}_{\bar{\rho}}(\mu_\ell)$$

[3] More strictly speaking, there exists a C^2 mapping $\Phi : V \to \mathbb{R}^n$ such that $\Phi(V)$ and V are diffeomorphic and furthermore $h(\Phi(v), v) = 0$ $(v \in V)$ holds by re-ordering the variables in x if necessary.

for a sufficiently large ℓ. Let

$$\check{x}_{\mathcal{M}}(\mu_\ell) \in \underset{y\in\mathcal{M}}{\arg\min} \|\check{x}(\mu_\ell) - y\|,$$
$$v_{\mu_\ell} := \Phi^{-1}(\check{x}_{\mathcal{M}}(\mu_\ell)),\ V_\ell := \left\{ v \in V \,\middle|\, \|v - v_{\mu_\ell}\| < \frac{\bar{\rho}\mu_\ell}{2M_1} \right\} \tag{3.27}$$

for each ℓ, where $\check{x}(\cdot)$ is defined in Lemma 2. Then, by (3.24),

$$\text{dist}\,(\check{x}(\mu_\ell), \mathcal{M}) = \|\check{x}_{\mathcal{M}}(\mu_\ell) - \check{x}(\mu_\ell)\| = \mathrm{O}(\mu_\ell^2). \tag{3.28}$$

For any $v \in V_\ell$,

$$\begin{aligned}\|\Phi(u) - \check{x}(\mu_\ell)\| &\leq \|\Phi(v) - \check{x}_{\mathcal{M}}(\mu_\ell)\| + \|\check{x}_{\mathcal{M}}(\mu_\ell) - \check{x}(\mu_\ell)\| \\ &\leq \|\Phi(v) - \Phi(v_{\mu_\ell})\| + \mathrm{O}(\mu_\ell^2) \\ &\leq M_1\|v - v_{\mu_\ell}\| + \mathrm{O}(\mu_\ell^2) \\ &\leq \frac{\bar{\rho}\mu_\ell}{2} + \mathrm{O}(\mu_\ell^2),\end{aligned}$$

where the second inequality follows from (3.27) and (3.28), the third one from (3.22), and the fourth from (3.27) and $v \in V_\ell$. Hence, we have $\|\Phi(v) - \check{x}(\mu)\| \leq \bar{\rho}\mu_\ell,\ \forall v \in V_\ell$ holds for sufficiently large ℓ, which yields that

$$\Phi(V_\ell) \subseteq \mathcal{P}_{\bar{\rho}}(\mu_\ell). \tag{3.29}$$

Furthermore, since both x^ℓ and y^ℓ converge to x^* as ℓ tends to ∞, Theorem 2 implies

$$\begin{aligned}\|\check{x}(\mu_\ell) - x^\ell\| &= \|x^* + \mu_\ell\xi^* - x^\ell\| \\ &= \mu_\ell\left\|\xi^* - \frac{x^\ell - x^*}{\mu_\ell}\right\| \\ &= o(\mu_\ell),\end{aligned}$$

and also $\|\check{x}(\mu_\ell) - y^\ell\| = o(\mu_\ell)$ in a similar way. These relations along with the triangle inequality and (3.28) yield

$$\max\left(\|\check{x}_{\mathcal{M}}(\mu_\ell) - x^\ell\|, \|\check{x}_{\mathcal{M}}(\mu_\ell) - y^\ell\|\right) = o(\mu_\ell),$$

which implies that for sufficiently large $\ell \geq 0$, $\max\left(\|\check{x}_{\mathcal{M}}(\mu_\ell) - x^\ell\|, \|\check{x}_{\mathcal{M}}(\mu_\ell) - y^\ell\|\right) \leq \frac{\mu_\ell\bar{\rho}}{4M_1M_2}$, thus

$$\|v_{\mu_\ell} - v^\ell\| = \|\Phi^{-1}(\check{x}_{\mathcal{M}}(\mu_\ell)) - \Phi^{-1}(x^\ell)\| \leq M_2\|\check{x}_{\mathcal{M}}(\mu_\ell) - x^\ell\| \leq \frac{\bar{\rho}\mu_\ell}{4M_1}.$$

Therefore, we gain $v^\ell \in \text{int}\, V_\ell$ for ℓ large enough. In a similar way, we can show $\theta^\ell \in \text{int}\, V_\ell$. In short, from the above arguments we obtain that

$$\{v^\ell, \theta^\ell\} \subseteq \text{int}\, V_\ell \tag{3.30}$$

for sufficiently large ℓ.

By letting $Y := \mu G(x)^{-1}$, (3.21) in Proposition 7 is translated as

$$d_v^\top \nabla^2\Psi_\mu(v) d_v > 0,\ \forall d_v \in \mathbb{R}^{n-s} \setminus \{0\}, \tag{3.31}$$

for any $v \in \Phi^{-1}(\Phi(V) \cap \mathcal{P}_\rho(\mu))$. Since inequality (3.31) holds for any $v \in V_\ell$ and V_ℓ is convex, the function Ψ_{μ_ℓ} is regarded as a strictly convex function in $\text{int}\, V_\ell$. Therefore, a point $v \in \text{int}\, V_\ell$ which fulfills $\nabla\Psi_{\mu_\ell}(v) = 0$ must be unique, which together with (3.29), (3.26), and (3.30) implies $\theta^\ell = v^\ell$. This gives $x^\ell = y^\ell$, a contradiction. With this, we ensure that by setting $\bar{\mu}$ to be small enough, a BKKT point $x(\mu)$ exists uniquely in $\Phi(V) \cap \text{int} \bigcup_{\mu\in(0,\mu_\ell]} \mathcal{P}_{\bar{\rho}}(\mu)$. We thus obtain the desired assertion.

Finally, the strict local optimality of $x(\mu)$ in problem (3.25) follows immediately from (3.31). The whole proof is complete. □

3.7. Main claim: existence and uniqueness of central path

Recall that $w^{\mathrm{a}} = (x^*, Y_{\mathrm{a}}, z^{\mathrm{a}})$ as defined in (3.2), and also recall the definitions of the sequence $\{x^k\}$ and the barrier parameter sequence $\{\mu_k\}$ organized in subsections 3.3 and 3.4. For each k, x^k is a BKKT point with barrier parameter μ_k and $\lim_{k\to\infty} x^k = x^*$. Theorem 3 implies that, given $\{\mu_k\}$, such x^k is uniquely determined for each k large enough. Taking this fact into consideration, we establish the following main result.

Theorem 4. *Suppose that Assumption 1 holds. There exist some $\bar{\mu} > 0$ and a "unique" central path $w(\mu) := (x(\mu), Y(\mu), z(\mu)) \in \mathcal{W}_{++}$ such that*

1. *$w(\mu)$ is smooth at any $\mu \in (0,\bar{\mu}]$ and $x(\mu)$ is a strict local optimum of problem (3.25) for each μ;*
2. $\lim\limits_{\mu\to 0+} w(\mu) = w^{\mathrm{a}}$.

Proof. Recall the definition of Ψ_μ in (3.25). Choose $\bar{\mu},\bar{\rho} > 0$ small enough. According to Theorem 3, for any barrier parameter $\mu \in (0,\bar{\mu}]$, a BKKT point, written x_μ, exists uniquely in $\Phi(V) \cap \operatorname{int} \bigcup_{\mu\in(0,\bar{\mu}]} \mathcal{P}_{\bar{\rho}}(\mu)$, which implies that

$$\nabla\Psi_\mu(v_\mu) = 0 \text{ with } v_\mu := \Phi^{-1}(x_\mu) \in V,\ \forall \mu \in (0,\bar{\mu}].$$

Moreover, $\nabla^2\Psi(v_\mu)$ is nonsingular for any $\mu \in (0,\bar{\mu}]$ since (3.31) holds with $v := v_\mu$. Hence, $\nabla^2\Psi(\bar{\mu}/2)$ is nonsingular from $\bar{\mu}/2 \in (0,\bar{\mu}]$. Then, by applying the implicit function theorem to $\nabla\Psi_\mu(v) = 0$ at $v = v_{\bar{\mu}/2}$, there exist $\delta_\mu \in (0,\bar{\mu}/3)$ and a smooth curve $v(\cdot) : [\bar{\mu}/2 - \delta_\mu, \bar{\mu}/2 + \delta_\mu] \to V$ such that

$$\nabla\Psi_\mu\,(v(\mu)) = 0,\ \forall \mu \in \left[\frac{\bar{\mu}}{2} - \delta_\mu, \frac{\bar{\mu}}{2} + \delta_\mu\right],\ v\left(\frac{\bar{\mu}}{2}\right) = v_{\frac{\bar{\mu}}{2}}.$$

By letting

$$x(\mu) := \Phi(v(\mu)) \text{ for each } \mu \in \left[\frac{\bar{\mu}}{2} - \delta_\mu, \frac{\bar{\mu}}{2} + \delta_\mu\right],$$

$x(\mu)$ is a BKKT point for each $\mu \in [\frac{\bar{\mu}}{2} - \delta_\mu, \frac{\bar{\mu}}{2} + \delta_\mu]$. By recalling the argument at the beginning of this proof and noting $\delta_\mu \in (0,\bar{\mu}/3)$, $x(\mu)$ must be unique in $\Phi(V) \cap \operatorname{int} \bigcup_{\mu\in(0,\bar{\mu}]} \mathcal{P}_{\bar{\rho}}(\mu)$ for each $\mu \in [\frac{\bar{\mu}}{2} - \delta_\mu, \frac{\bar{\mu}}{2} + \delta_\mu] \subseteq (0,\bar{\mu}]$. Then, by repeating the above argument at the end-point of the curve $x(\mu)$, i.e., the point corresponding to $\mu = \frac{\bar{\mu}}{2} - \delta_\mu$, it can be smoothly extended towards $\mu = 0$. Namely,

$$\inf\left\{ a > 0 \;\middle|\; x(\cdot) \text{ is defined in } \left[a, \tfrac{\bar{\mu}}{2} + \delta_\mu\right] \right\} = 0. \tag{3.32}$$

Therefore, the smooth curve $x(\cdot)$ can be defined in $(0,\bar{\mu}/2]$ and moreover, such $x(\cdot)$ is determined uniquely in $\Phi(V) \cap \bigcup_{\mu\in(0,\bar{\mu}]} \mathcal{P}_{\bar{\rho}}(\mu)$ by Theorem 3 again. In fact, any sequence of BKKT points which converges to x^* must be contained eventually in $\Phi(V) \cap \bigcup_{\mu\in(0,\bar{\mu}]} \mathcal{P}_{\bar{\rho}}(\mu)$.[4] Hence, we conclude that $x(\cdot)$ such that $\lim_{\mu\to 0+} x(\mu) = x^*$ is unique in the whole space.

Next, note that $\nabla f(x(\mu)) - \mathcal{J}G(x(\mu))^* \mu G(x(\mu))^{-1} \in \operatorname{Im}\nabla h\,(x(\mu))$ because $x(\mu)$ is a BKKT point with barrier parameter μ. Then, letting

$$Y(\mu) := \mu G(x(\mu))^{-1},$$
$$z(\mu) := -(\nabla h(x(\mu))^\top \nabla h(x(\mu)))^{-1} \nabla h(x(\mu))^\top \left(\nabla f(x(\mu)) - \mathcal{J}G(x(\mu))^* Y(\mu)\right)$$

for $\mu > 0$, we ensure $w(\mu) := (x(\mu), Y(\mu), z(\mu))$ is a BKKT triplet with barrier parameter $\mu > 0$ and $w(\cdot)$ is a unique smooth path such that $\lim_{\mu\to 0+} x(\mu) = x^*$. Thus, item 1 of this theorem is obtained. Moreover, $\lim_{\mu\to 0+} w(\mu) = w^{\mathrm{a}}$ follows from Theorem 1. Hence, item 2 is also obtained. □

[4] This fact is verified because Corollary 1 yields that the distance between a sequence of BKKT points converging to x^* and the line $x^* + s\frac{\xi^*}{\|\xi^*\|}$ $(s \geq 0)$ converges to 0.

4. Concluding remarks and future work

In this paper, we have studied properties of a central path for nonlinear semidefinite optimization problems (NSDPs). Specifically, we have proven that, under the strict complementarity condition, strong second-order sufficient condition, and Mangasarian-Fromovitz constraint qualification, there exists a smooth central path which converges to a KKT triplet with an analytic center. In particular, given a KKT triplet, a central path leading to that KKT triplet is uniquely determined. Unlike the past results concerning the central path for the NSDP, the nondegeneracy condition is not assumed there. The author believes that the results obtained in this paper will play substantial role for further development of the primal-dual interior point method for the NSDP.

There exist two directions for future works. The first one is concerned with limiting behavior of the tangential direction $\dot{x}(\mu)$ in the x-space as $\mu \to 0$. We have the following conjecture:

$$\lim_{\mu\to 0+} \dot{x}(\mu) = \xi^*,$$

where ξ^* was defined in Proposition 5. For nonlinear optimization, the corresponding result was proven by Wright and Orban [58, Theorem 12]. The second direction of future works is to mitigate the strict complementarity (SC) condition from our assumptions. The SC actually plays a key role in our analysis, particularly when establishing Proposition 3 that is a base for proving the subsequent theorems.

A. Omitted Proofs

In this appendix, we give the proofs which are not shown in the main part of this paper.

A.1. Proof of Lemma 1

The second equality follows from the direct calculation along with the fact of $\Delta G^{\mathrm{EF}} = (\Delta G^{\mathrm{FE}})^\top$, and the first one is derived from the following transformation:

$$\begin{aligned}
& d^\top \Omega(x^*, Y_*) d \\
=& 2\mathrm{Tr}\left(\sum_{j=1}^n \sum_{i=1}^n d_i d_j Y_* \mathcal{G}_i^* G_*^\dagger \mathcal{G}_j^*\right) \\
=& 2\mathrm{Tr}\left((P_*^\top Y_* P_*)\left(\sum_{i=1}^n d_i P_*^\top \mathcal{G}_i^* P_*\right)\left(P_*^\top G_*^\dagger P_*\right)\left(\sum_{j=1}^n d_j P_*^\top \mathcal{G}_j^* P_*\right)\right) \\
=& 2\mathrm{Tr}\left(\begin{bmatrix} Y_*^{\mathrm{EE}} & O \\ O & O \end{bmatrix}\begin{bmatrix} \Delta G^{\mathrm{EE}}(x^*; d) & \Delta G^{\mathrm{EF}}(x^*; d) \\ \Delta G^{\mathrm{FE}}(x^*; d) & \Delta G^{\mathrm{FF}}(x^*; d) \end{bmatrix}\begin{bmatrix} O & O \\ O & (G_*^{\mathrm{FF}})^{-1} \end{bmatrix}\begin{bmatrix} \Delta G^{\mathrm{EE}}(x^*; d) & \Delta G^{\mathrm{EF}}(x^*; d) \\ \Delta G^{\mathrm{FE}}(x^*; d) & \Delta G^{\mathrm{FF}}(x^*; d) \end{bmatrix}\right) \\
=& 2\mathrm{Tr}\left(Y_*^{\mathrm{EE}} \Delta G^{\mathrm{FE}}(x^*; d)(G_*^{\mathrm{FF}})^{-1} \Delta G^{\mathrm{EF}}(x^*; d)\right).
\end{aligned}$$

The proof is complete. □

A.2. Proof of Proposition 1

We first show the first assertion: the unique existence of the analytic center at x^*. Note that because of relation (2.8) for $(Y, z) \in \Lambda(x^*)$, (3.3) is equivalent to the following problem with respect to only Y:

$$\begin{aligned}
\min_{Y \in \mathbb{S}^m} \quad & -\log\det Y^{\mathrm{EE}} \\
\text{s.t.} \quad & \nabla f(x^*) - \mathcal{J} G^{\mathrm{EE}}(x^*)^* Y^{\mathrm{EE}} \in \mathrm{Im} \nabla h(x^*), \\
& Y^{\mathrm{EF}} = Y^{\mathrm{FE}} = O, \ Y^{\mathrm{FF}} = O, \\
& Y^{\mathrm{EE}} \in \mathbb{S}_+^{m - r_*},
\end{aligned} \tag{A.1}$$

where $\mathcal{J} G^{\mathrm{EE}}(x^*)^* Z := [(E_*^\top \mathcal{G}_i(x^*) E_*) \bullet Z]_{i=1}^n \in \mathbb{R}^n$ for $Z \in \mathbb{S}^{m-r_*}$.

We establish existence of optima of (3.3). By the strict complementarity condition as for the NSDP, there exists $(Y, z) \in \Lambda(x^*)$ such that $Y + G_* \in \mathbb{S}_{++}^m$, which implies $Y^{\mathrm{EE}} \in \mathbb{S}_{++}^{m-r_*}$. This means that a finite objective value of (A.1) is attained at such a matrix Y. Moreover, as $\Lambda(x^*)$ is convex and bounded from the MFCQ at x^* for the NSDP, so is the feasible region of (A.1). By combining these facts, (A.1) is ensured to have an optimum, say $Y_a \in \mathbb{S}^m_+$. From the full column rankness of $\nabla h(x^*)$, we see that the linear equation $\nabla f(x^*) - \mathcal{J}G^{\mathrm{EE}}(x^*)^* Y_a^{\mathrm{EE}} + \nabla h(x^*) z = 0$ has a unique solution $z \in \mathbb{R}^s$, written z^a. This (Y_a, z^a) is nothing but an optimum of (3.3).

Next, consider the following problem:

$$\begin{array}{ll} \min\limits_{Z} & -\log\det Z \\ \text{s.t.} & \nabla f(x^*) - \mathcal{J}G^{\mathrm{EE}}(x^*)^* Z \in \mathrm{Im}\nabla h(x^*), \\ & Z \in \mathbb{S}^{m-r_*}_+. \end{array} \tag{A.2}$$

For a feasible point Y of (A.1), Y^{EE} is clearly feasible to (A.2), and hence so is Y_a^{EE} to (A.2). Furthermore, we can ensure that Y_a^{EE} is optimal to (A.2). Indeed, if not, there exists Z such that Z is feasible to (A.2) and $-\log\det Z < -\log\det Y_a^{\mathrm{EE}}$. Since $Y := E_* Z E_*^\top \in \mathbb{S}^m_+$ is feasible to (A.1) and $\log\det Y^{\mathrm{EE}} = \log\det Z$, we gain $-\log\det Y^{\mathrm{EE}} < -\log\det Y_a^{\mathrm{EE}}$, a contradiction to the optimality of Y_a^{EE} for (A.1). Lastly, since (A.2) is a strictly convex problem, we see that Y_a^{EE} is a *unique* optimum of (A.2).

In turn, we establish the uniqueness of $(Y_a, z^a) \in \mathbb{S}^m_+ \times \mathbb{R}^\ell$ as optimum of (3.3). To derive a contradiction, assume that there exist two distinct optima (Y_a, z^a) and $(\widetilde{Y}_a, \widetilde{z^a})$ at x^*, which yields that Y_a and $\widetilde{Y}_a$ are both optima of (A.1) by the preceding argument. Thus, so are Y_a^{EE} and $\widetilde{Y}_a^{\mathrm{EE}}$ to (A.2), in particular $Y_a^{\mathrm{EE}} = \widetilde{Y}_a^{\mathrm{EE}}$, according to the preceding argument again. Hence, we have

$$P_*^\top (Y_a - \widetilde{Y}_a) P_* = \begin{bmatrix} Y_a^{\mathrm{EE}} - \widetilde{Y}_a^{\mathrm{EE}} & O \\ O & O \end{bmatrix} = O.$$

Since P_* is nonsingular, we obtain $Y_a = \widetilde{Y}_a$, which together with the full column rankness of $\nabla h(x^*)$ implies $z^a = \widetilde{z^a}$. Hence we ensure $(Y_a, z^a) = (\widetilde{Y}_a, \widetilde{z^a})$, which is a contradiction. Consequently, (3.3) has a unique optimum, and thus we obtain the first claim.

There remains to verify the second claim as for (3.4). If (Y_a, z^a) is the analytic center at x^*, $(Y_a, z^a) \in \Lambda(x^*)$ holds by definition, and from the above proof, Y_a^{EE} is the unique optimum of (A.2). Hence, by the KKT conditions of (A.2), there exists $v \in \mathbb{R}^s$ such that (3.4) holds. Conversely, if such v exists and $(Y_a, z^a) \in \Lambda(x^*)$, Y_a^{EE} solves (A.2), and hence $E_* Y_a^{\mathrm{EE}} E_*^\top = Y_a$ does (A.1). This means that (Y_a, z^a) is the analytic center. The whole proof is complete. □

A.3. Proof of Proposition 2

Take a compact set $B \subseteq \mathbb{R}^n$ with nonempty interior such that $x^* \in \operatorname{int} B$ and it is a unique optimum of the problem

$$\min\ f(x) \text{ s.t. } h(x) = 0,\ G(x) \in \mathbb{S}^m_+,\ x \in B. \tag{A.3}$$

Consider the sequence of the relevant barrier problems parameterized with μ_k as in the following:

$$\min\ f(x) - \mu_k \log\det G(x) \text{ s.t. } h(x) = 0,\ G(x) \in \mathbb{S}^m_{++},\ x \in B, \tag{A.4}$$

and let x^k be an optimum of problem (A.4) for each k.

We will prove the theorem by showing that the above-defined sequence $\{x^k\}$ is nothing but the desired one. To this end, it suffices to prove that $\{x^k\}$ converges to x^*. Indeed, because $x^* \in \operatorname{int} B$, the constraint $x \in B$ for problem (A.4) is inactive at x^k for sufficiently large k, and thus x^k eventually becomes a local optimum of (3.5).

We consider the first case (i) where $G(x^*) \in \mathbb{S}^m_+ \setminus \mathbb{S}^m_{++}$, i.e., $G(x^*)$ is on the boundary of $\mathbb{S}^m_+$ and thus $\det G(x^*) = 0$. The proof for the other case (ii) where $G(x^*) \in \mathbb{S}^m_{++}$ will be given later. Letting $\varphi_k := f(x^k) - \mu_k \log\det G(x^k)$ for each k, the first goal is to prove

$$\lim_{k\to\infty} \varphi_k = f(x^*). \tag{A.5}$$

Without loss of generality, by re-taking a smaller B with $\operatorname{int} B \ni x^*$ if necessary, we can suppose that $\det G(x) < 1$ for all $x \in B$ because of $\det G(x^*) = 0$, yielding

$$-\log\det G(x) > 0, \quad \forall x \in B, \tag{A.6}$$

which together with the feasibility of x^k for (A.3) implies

$$-\mu_k \log\det G(x^k) > 0 > f(x^*) - f(x^k). \tag{A.7}$$

Using the two inequalities in (A.7) yields

$$\begin{aligned} f(x^*) &< f(x^k) \\ &< f(x^k) - \mu_k \log\det G(x^k)\ (= \varphi_k) \\ &\leq f(x^{k-1}) - \mu_k \log\det G(x^{k-1}) \\ &\leq f(x^{k-1}) - \mu_{k-1} \log\det G(x^{k-1})\ (= \varphi_{k-1}), \end{aligned} \tag{A.8}$$

where the third inequality follows from the optimality of x^k for problem (A.4) and the fourth one is due to $\mu_k \leq \mu_{k-1}$ and $-\log\det G(x^{k-1}) > 0$ from (A.6) and $x^{k-1} \in B$. From the above inequalities, we find that $\{\varphi_k\}$ is a nonincreasing sequence such that it is bounded by $f(x^*)$ from below. Therefore, we ensure the existence of $\lim_{k\to\infty} \varphi_k$ and moreover obtain

$$f(x^*) \leq \lim_{k\to\infty} \varphi_k. \tag{A.9}$$

To verify (A.5), there remains to prove the converse inequality. Related to $\{x^k\}$, under the MFCQ at x^*. we can pick another sequence $\{x^{\ell(k)}\}$ feasible to problem (A.3) such that it converges to x^* and also satisfies $\det G(x^{\ell(k)}) = \mu_k$ for each $k \geq K$ with sufficiently large $K > 0$.[5] We then obtain $\lim_{k\to\infty} \varphi_k \leq f(x^*)$ since $x \log x \to 0$ as $x \to 0+$ and $\varphi_k \leq f(x^{\ell(k)}) - \mu_k \log\det G(x^{\ell(k)})$ holds by the definition of x^k. Together with (A.9), it derives the target equation (A.5).

The convergence of $\{x^k\}$ to x^* is not difficult to derive from (A.5). Letting $\bar{x}$ be an arbitrary accumulation point of $\{x^k\}$ and taking into consideration $-\mu_k \log\det G(x^k) > 0$ in φ_k, we get $\limsup_{k\to\infty} \varphi_k \geq f(\bar{x})$, which combined with (A.5) implies $f(x^*) \geq f(\bar{x})$. By the feasibility of $\bar{x}$ and the unique optimality of x^* for (A.3), we gain $x^* = \bar{x}$. Finally, since $\bar{x}$ was an arbitrary accumulation point of $\{x^k\}$, we conclude that $\lim_{k\to\infty} x^k = x^*$.

We next consider case (ii) where $G(x^*) \in \mathbb{S}^m_{++}$. Note that $\log\det G(x^*)$ is finite in this case. Without loss of generality, we may assume that $\det G(x) > 0$ for all $x \in B$, by taking a smaller $B (\ni x^*)$ if necessary. Let $\bar{x}$ be an arbitrary accumulation point of $\{x^k\}$ and note that $\log\det G(\bar{x})$ is also finite since $\det G(\bar{x}) > 0$ by virtue of $\bar{x} \in B$. By the optimality of x^k and feasibility of x^* to (A.4), it follows that

$$f(x^k) - \mu_k \log\det G(x^k) \leq f(x^*) - \mu_k \log\det G(x^*),$$

where driving $k \to \infty$ and taking a subsequence if necessary imply $f(\bar{x}) \leq f(x^*)$. Then, in virtue of feasibility of $\bar{x}$ and unique optimality of x^* for (A.3), we have $x^* = \bar{x}$. We hence conclude that $\lim_{k\to\infty} x^k = x^*$ as for case(ii).

Consequently, the desired result is obtained and the proof is complete. □

[5] This fact is verified as follows: From the MFCQ at x^*, there exists $d \in \mathbb{R}^n$ such that $G(x^*) + \Delta G(x; d) \in \mathbb{S}^m_{++}$ and $\nabla h(x^*)^\top d = 0$. By the full column rankness of $\nabla h(x^*)$, we can ensure existence of a smooth curve $x(\cdot) : [0, \bar{t}] \to \mathbb{R}^s$ with some $\bar{t} > 0$ such that $x(0) = x^*$, $\dot{x}(0) = d$, and $x(t)$ is feasible to (A.4), $\forall t \in (0, \bar{t}]$. Particularly, $G(x(t)) \in \mathbb{S}^m_{++}$ holds for all $t \in (0, \bar{t}]$. Therefore, as $\det G(x(t))$ is continuous w.r.t. $t \geq 0$ and takes 0 at $t = 0$ by the assumption $G(x^*) \in \mathbb{S}^m_+ \setminus \mathbb{S}^m_{++}$, we conclude that for any sufficiently small $\alpha > 0$, $\det G(x(t)) = \alpha$ is attained by some $t \in (0, \bar{t}]$. The proof is complete.

A.4. Proof of Proposition 3

We use the notations described before Proposition 3. In particular, recall (2.10) and (2.11).

Proof of Proposition 3: To begin with, for each $k \geq 0$, let

$$\widetilde{d}^k := \frac{d^k}{\|d^k\|}.$$

Since $\{\widetilde{d}^k\}$ is bounded, it has at least one accumulation point, say $\widetilde{d}^*$. Choose an arbitrary subsequence $\{\widetilde{d}^k\}_{k\in\mathcal{K}}$ which converges to $\widetilde{d}^*$. From remark 3, $\{w^k\}_{k\in\mathcal{K}}$ has an accumulation point, say $w^* := (x^*, Y_*, z^*)$. Without loss of generality, we assume $\lim_{\mathcal{K}\ni k\to\infty} w^k = w^*$.

We prove the assertion by two steps. As the first step, we prove

$$\liminf_{k\to\infty} \frac{\mu_k}{\|d^k\|} > 0. \tag{A.10}$$

In order to derive a contradiction, suppose to the contrary that there exists a subsequence of $\left\{\frac{\mu_k}{\|d^k\|}\right\}_{k\in\mathcal{K}}$ such that it converges to 0. We may assume $\lim_{k(\in\mathcal{K})\to\infty} \frac{\mu_k}{\|d^k\|} = 0$ by retaking $\mathcal{K}$ if necessary. Since $\lim_{k(\in\mathcal{K})\to\infty} \widetilde{d}^k = \widetilde{d}^*$, $\widetilde{d}^*$ satisfies

$$\left(E_*^\top \Delta G(x^*;\widetilde{d}^*)E_* =\right) \Delta G^{\mathrm{EE}}(x^*;\widetilde{d}^*) \in \mathbb{S}_+^{r_*},\ \nabla h(x^*)^\top \widetilde{d}^* = 0, \tag{A.11}$$

where these relations are derived from dividing the following equations by $\|d^k\|$ and passing to the limit:

$$\begin{aligned} \mathbb{S}_{++}^{r} \ni E_*^\top G_k E_* = E_*^\top (G_k - G_*) E_* &= \Delta G^{\mathrm{EE}}(x^*; d^k) + \mathrm{O}(\|d^k\|^2), \\ 0 = h(x^k) &= h(x^*) + \nabla h(x^*)^\top d^k + \mathrm{O}(\|d^k\|^2). \end{aligned}$$

As $w^k = (x^k, Y_k, z^k)$ satisfies the BKKT conditions and $P_* = [E_*, F_*]$ is an orthogonal matrix, we obtain, for each $k \in \mathcal{K}$,

$$\begin{aligned} \frac{\mu_k I_{r_*}}{\|d^k\|} &= \frac{E_*^\top G_k Y_k E_*}{\|d^k\|} \\ &= \frac{E_*^\top \left(G(x^*) + \Delta G(x^*; d^k) + \mathrm{O}(\|d^k\|^2)\right) \begin{bmatrix} E_* & F_* \end{bmatrix} \begin{bmatrix} E_*^\top \\ F_*^\top \end{bmatrix} Y_k E_*}{\|d^k\|}, \end{aligned}$$

which together with driving $k(\in\mathcal{K}) \to \infty$ yields

$$\Delta G^{\mathrm{EE}}(x^*;\widetilde{d}^*) Y_*^{\mathrm{EE}} = O, \tag{A.12}$$

where we have used the relations $G_*^{\mathrm{EE}} = O$ and $Y_*^{\mathrm{FE}} = O$ from (2.9).

As $w^* = (x^*, Y_*, z^*)$ and $(x^*, Y_{\mathrm{a}}, z^{\mathrm{a}})$ satisfy the KKT conditions, it follows that

$$\begin{aligned} \nabla f(x^*) &= \mathcal{J}G(x^*)^* Y_* - \nabla h(x^*) z^*, && \text{(A.13)} \\ &= \mathcal{J}G(x^*)^* Y_{\mathrm{a}} - \nabla h(x^*) z^{\mathrm{a}}. && \text{(A.14)} \end{aligned}$$

Pre-multiplying both (A.13) and (A.14) by $(\widetilde{d}^*)^\top$ and noting (A.12) lead to

$$\nabla f(x^*)^\top \widetilde{d}^* = \mathrm{Tr}\left(\Delta G^{\mathrm{EE}}(x^*;\widetilde{d}^*) Y_{\mathrm{a}}^{\mathrm{EE}}\right) = \mathrm{Tr}\left(\Delta G^{\mathrm{EE}}(x^*;\widetilde{d}^*) Y_*^{\mathrm{EE}}\right) = 0. \tag{A.15}$$

From (A.15) and (A.11), we ensure

$$\widetilde{d}^* \in C(x^*), \tag{A.16}$$

where $C(x^*)$ is defined in (2.13). As $Y_{\mathrm{a}}^{\mathrm{EE}}$ is positive definite by definition and $\Delta G^{\mathrm{EE}}(x^*;\widetilde{d}^*) \in \mathbb{S}_+^m$ follows from (A.11) again, $\mathrm{Tr}\left(\Delta G^{\mathrm{EE}}(x^*;\widetilde{d}^*) Y_{\mathrm{a}}^{\mathrm{EE}}\right) = 0$ in (A.15) yields

$$\Delta G^{\mathrm{EE}}(x^*;\widetilde{d}^*) = O. \tag{A.17}$$

Next, we transform $(\widetilde{d^*})^\top \mathcal{J}G(x^*)^*(Y_k - \overline{Y})$ as

$$\begin{aligned}(\widetilde{d^*})^\top \mathcal{J}G(x^*)^*(Y_k - Y_*) &= \Delta G(x^*;\widetilde{d^*}) \bullet (Y_k - Y_*)\\ &= \mathrm{Tr}\left(\begin{bmatrix}E_*^\top\\ F_*^\top\end{bmatrix}\Delta G(x^*;\widetilde{d^*})\begin{bmatrix}E_* & F_*\end{bmatrix}\begin{bmatrix}E_*^\top\\ F_*^\top\end{bmatrix}(Y_k - Y_*)\begin{bmatrix}E_* & F_*\end{bmatrix}\right)\\ &= \begin{bmatrix}\Delta G^{\mathrm{EE}}(x^*;\widetilde{d^*}) & \Delta G^{\mathrm{EF}}(x^*;\widetilde{d^*})\\ \Delta G^{\mathrm{FE}}(x^*;\widetilde{d^*}) & \Delta G^{\mathrm{FF}}(x^*;\widetilde{d^*})\end{bmatrix} \bullet \begin{bmatrix}Y_k^{\mathrm{UU}} - Y_*^{\mathrm{FF}} & Y_k^{\mathrm{UN}}\\ Y_k^{\mathrm{NU}} & Y_k^{\mathrm{NN}}\end{bmatrix}\\ &= \begin{bmatrix}O & \Delta G^{\mathrm{EF}}(x^*;\widetilde{d^*})\\ \Delta G^{\mathrm{FE}}(x^*;\widetilde{d^*}) & \Delta G^{\mathrm{FF}}(x^*;\widetilde{d^*})\end{bmatrix} \bullet \begin{bmatrix}Y_k^{\mathrm{UU}} - Y_*^{\mathrm{FF}} & Y_k^{\mathrm{UN}}\\ Y_k^{\mathrm{NU}} & Y_k^{\mathrm{NN}}\end{bmatrix}\\ &= 2\mathrm{Tr}\left(\Delta G^{\mathrm{EF}}(x^*;\widetilde{d^*})Y_k^{\mathrm{NU}}\right) + \mathrm{Tr}\left(\Delta G^{\mathrm{FF}}(x^*;\widetilde{d^*})Y_k^{\mathrm{NN}}\right),\end{aligned} \tag{A.18}$$

where the second equality follows from the fact that $[E_*, F_*](= P_*)$ is an orthogonal matrix and the fourth one is due to (A.17).

Since w^k and w^* satisfy the BKKT and KKT conditions, respectively, we have $\nabla_x L(w^*) = \nabla_x L(w^k) = 0$ for each $k \in \mathcal{K}$, yielding

$$\begin{aligned}0 &= (\widetilde{d^*})^\top \frac{(\nabla_x L(w^k) - \nabla_x L(w^*))}{\|d^k\|}\\ &= (\widetilde{d^*})^\top \frac{\nabla^2_{xx} L(w^*)(d^k) - \mathcal{J}G(x^*)(Y_k - Y_*) + \nabla h(x^*)^\top (z^k - z^*) + \mathrm{O}(\|d^k\|^2)}{\|d^k\|}\\ &= \frac{(\widetilde{d^*})^\top \nabla^2_{xx} L(w^*)(d^k) - \Delta G(x^*;\widetilde{d^*}) \bullet (Y_k - Y_*) + \mathrm{O}(\|d^k\|^2)}{\|d^k\|}\\ &= (\widetilde{d^*})^\top \nabla^2_{xx} L(w^*)\frac{d^k}{\|d^k\|} - \frac{2\mathrm{Tr}\left(\Delta G^{\mathrm{EF}}(x^*;\widetilde{d^*})Y_k^{\mathrm{NU}}\right) + \mathrm{Tr}\left(\Delta G^{\mathrm{FF}}(x^*;\widetilde{d^*})Y_k^{\mathrm{NN}}\right)}{\|d^k\|} + \mathrm{O}(\|d^k\|),\end{aligned} \tag{A.19}$$

where the last equality follows from (A.18). Notice that the off-diagonal elements of $P_*^\top G_k Y_k P_* (= \mu_k I)$ are zeros for all k. Hence, for each $k (\in \mathcal{K}) \geq 0$, we have

$$\begin{aligned}O &= F_*^\top G_k Y_k E_*\\ &= F_*^\top G_k \begin{bmatrix}E_*, F_*\end{bmatrix}\begin{bmatrix}E_*^\top\\ F_*^\top\end{bmatrix} Y_k E_*\\ &= F_*^\top G_k E_* Y_k^{\mathrm{EE}} + F_*^\top G_k F_* Y_k^{\mathrm{FE}}\end{aligned}$$

for each $k \in \mathcal{K}$. Substituting Taylor's expansion $G_k = G_* + \Delta G(x^*; d^k) + \mathrm{O}(\|d^k\|^2)$ into the last equation yields

$$(G_*^{\mathrm{FE}} + \Delta G^{\mathrm{FE}}(x^*; d^k))Y_k^{\mathrm{EE}} + \left(G_*^{\mathrm{FF}} + \Delta G^{\mathrm{FF}}(x^*; d^k)\right) Y_k^{\mathrm{FE}} = \mathrm{O}(\|d^k\|^2), \tag{A.20}$$

where $\|Y_k\|_{\mathrm{F}} = \mathrm{O}(1)$ was used for the last equality. Noting $G_*^{\mathrm{FE}} = O$ and dividing both the sides of the above by $\|d^k\|$ give

$$\frac{\Delta G^{\mathrm{FE}}(x^*; d^k)}{\|d^k\|} Y_k^{\mathrm{EE}} + \left(G_*^{\mathrm{FF}} \frac{Y_k^{\mathrm{FE}}}{\|d^k\|} + \frac{\Delta G^{\mathrm{FF}}(x^*; d^k)}{\|d^k\|} Y_k^{\mathrm{FE}}\right) = \mathrm{O}(\|d^k\|) \tag{A.21}$$

for $k \in \mathcal{K}$. Note that $\lim_{k\to\infty} Y_k^{\mathrm{FE}} = O$ holds, which implies $\lim_{k\to\infty} \frac{\Delta G^{\mathrm{FF}}(x^*;d^k)}{\|d^k\|} Y_k^{\mathrm{FE}} = O$. Moreover, together with letting $k (\in \mathcal{K}) \to \infty$, equation (A.21) implies

$$\lim_{k\to\infty} \frac{Y_k^{\mathrm{FE}}}{\|d^k\|} = -(G_*^{\mathrm{FF}})^{-1} \Delta G^{\mathrm{FE}}(x^*; \widetilde{d^*}) Y_*^{\mathrm{EE}}. \tag{A.22}$$

On the other hand, the $(2,2)$-block matrix of $P_*^\top G_k Y_k P_*/\|d^k\|(=\mu_k I/\|d^k\|)$ is calculated as

$$\begin{aligned}
&\frac{1}{\|d^k\|}\left(F_*^\top\left(G_*+\Delta G(x^k;d^k)\right)E_*Y_k^{\mathrm{EF}}+F_*^\top\left(G_*+\Delta G_k(x^k;d^k)\right)F_*Y_k^{\mathrm{FF}}\right)+\frac{\mathrm{O}(\|d^k\|^2)}{\|d^k\|}\\
=&\frac{1}{\|d^k\|}\left(F_*^\top\Delta G(x^k;d^k)E_*Y_k^{\mathrm{EF}}+\left(G_*^{\mathrm{FF}}+F_*^\top\Delta G_k(x^k;d^k)F_*\right)Y_k^{\mathrm{FF}}\right)+\mathrm{O}(\|d^k\|),
\end{aligned}\tag{A.23}$$

where we have used

$$G_*^{\mathrm{FE}}=O,\;G_*^{\mathrm{FF}}=O \tag{A.24}$$

from (2.9). In particular, from $\lim_{k\to\infty}\Delta G(x^k;d^k)/\|d^k\|=\Delta G(x^*;\widetilde{d^*})$ and $\lim_{k\to\infty}(Y_k^{\mathrm{EF}},Y_k^{\mathrm{FF}})=(O,O)$, we see

$$\lim_{k\to\infty}\frac{F_*^\top\Delta G(x^k;d^k)E_*Y_k^{\mathrm{EF}}+F_*^\top\Delta G(x^k;d^k)F_*Y_k^{\mathrm{FF}}}{\|d^k\|}=O.$$

Since the limit of (A.23) is zero by the assumption $\lim_{k\to\infty}\mu_k/\|d^k\|=0$ again and recalling that (A.23) is the $(2,2)$-block of $\mu_k I/\|d^k\|$, the above equation yields

$$\lim_{k\to\infty}\frac{G_*^{\mathrm{FF}}Y_k^{\mathrm{FF}}}{\|d^k\|}=O,$$

which combined with the nonsingularity of G_*^{FF} induces

$$\lim_{k\to\infty}\frac{Y_k^{\mathrm{FF}}}{\|d^k\|}=O. \tag{A.25}$$

By taking (A.22) and (A.25) into consideration and driving $k\to\infty$ in the equation in (A.19), it holds that

$$(\widetilde{d^*})^\top\nabla_{xx}^2L(w^*)\widetilde{d^*}=-2\mathrm{Tr}\left(Y_*^{\mathrm{EE}}\Delta G^{\mathrm{EF}}(x^*;\widetilde{d^*})(G_*^{\mathrm{FF}})^{-1}\Delta G^{\mathrm{FE}}(x^*;\widetilde{d^*})\right).$$

Combined with Lemma 1, this equation further implies

$$(\widetilde{d^*})^\top\nabla_{xx}^2L(w^*)\widetilde{d^*}+(\widetilde{d^*})^\top\Omega(x^*,Y_*)\widetilde{d^*}=0.$$

However, in view of $\widetilde{d^*}\in C(x^*)$ by (A.16) and $\widetilde{d^*}\neq 0$, the above equation contradicts the SSOSC. Therefore, we conclude (A.10).

In turn, we show $\frac{\mu_k}{\|d^k\|}=\mathrm{O}(1)$ as the second step. As w^k satisfies the BKKT conditions, we obtain, for each k,

$$\begin{aligned}
\frac{\mu_k I_{r_*}}{\|d^k\|}&=\frac{E_*^\top G_k Y_k E_*}{\|d^k\|}\\
&=\frac{E_*^\top\left(G_*+\mathcal{J}G(x^*)(d^k)+\mathrm{O}(\|d^k\|^2)\right)\begin{bmatrix}E_* & F_*\end{bmatrix}\begin{bmatrix}E_*^\top\\ F_*^\top\end{bmatrix}Y_kE_*}{\|d^k\|},
\end{aligned}$$

which together with (A.24) yields

$$\lim_{k\to\infty}\frac{\mu_k I_{r_*}}{\|d^k\|}=\Delta G^{\mathrm{EE}}(x^*;\widetilde{d^*})Y_*^{\mathrm{EE}}.$$

This means that the sequence $\left\{\frac{\mu_k}{\|d^k\|}\right\}$ is bounded, and we thus obtain the desired consequence. By combining (A.10) and this fact, the proof is complete. □

A.5. Proof of Proposition 5

To start with, decompose a vector $\Delta x \in \mathbb{R}^n$ into orthogonal component vectors as follows:

$$\Delta x = U_{x^*}\eta^1 + V_*\eta^2, \tag{A.26}$$

where $(\eta^1, \eta^2) \in \mathbb{R}^{p_*} \times \mathbb{R}^{n-p_*}$ and $V_* \in \mathbb{R}^{n\times(n-p_*)}$ is a matrix whose columns form an normal orthogonal basis of the orthogonal complement subspace for $\mathcal{U}_*$, where $\mathcal{U}_*$ is defined in (3.17).

Let $U^i_{x^*}$ be the i-th column of U_{x^*} for each $i = 1, 2, \ldots, p_*$. From (3.15), we have

$$\mathrm{Sym}\left(\begin{bmatrix} Y^{\mathrm{EE}}_{\mathrm{a}}\Delta G^{\mathrm{EE}}(x^*;\Delta x) & Y^{\mathrm{EE}}_{\mathrm{a}}\Delta G^{\mathrm{EF}}(x^*;\Delta x) \\ G^{\mathrm{FF}}_*\Delta Y^{\mathrm{FE}} & G^{\mathrm{FF}}_*\Delta Y^{\mathrm{FF}} \end{bmatrix}\right) = I,$$

of which the block components together with $G^{\mathrm{FF}}_* \in \mathbb{S}^{m-r_*}_{++}$ and $Y^{\mathrm{EE}}_{\mathrm{a}} \in \mathbb{S}^{r_*}_{++}$ yield

$$\Delta Y^{\mathrm{FF}} = (G^{\mathrm{FF}}_*)^{-1}, \tag{A.27}$$
$$\Delta G^{\mathrm{EE}}(x^*;\Delta x) = (Y^{\mathrm{EE}}_{\mathrm{a}})^{-1}, \tag{A.28}$$
$$Y^{\mathrm{EE}}_{\mathrm{a}}\Delta G^{\mathrm{EF}}(x^*;\Delta x) + \Delta Y^{\mathrm{EF}} G^{\mathrm{FF}}_* = O. \tag{A.29}$$

By (3.14), we obtain

$$U^\top_{x^*}\nabla^2_{xx}L(w^{\mathrm{a}})\Delta x - \left(\begin{bmatrix} O & \Delta G^{\mathrm{EF}}(x^*;U^i_{x^*}) \\ \Delta G^{\mathrm{FE}}(x^*;U^i_{x^*}) & \Delta G^{\mathrm{FF}}(x^*;U^i_{x^*}) \end{bmatrix} \bullet P^\top_*\Delta Y P_*\right)_{i=1}^{p_*} = 0,$$

leading to

$$U^\top_{x^*}\nabla^2_{xx}L(w^{\mathrm{a}})\Delta x - \left(\Delta G^{\mathrm{FE}}(x^*;U^i_{x^*})\bullet\Delta Y^{\mathrm{EF}} + \Delta G^{\mathrm{EF}}(x^*;U^i_{x^*})\bullet\Delta Y^{\mathrm{FE}} + \Delta G^{\mathrm{FF}}(x^*;U^i_{x^*})\bullet\Delta Y^{\mathrm{FF}}\right)_{i=1}^{p_*} = 0,$$

which is rephrased as

$$U^\top_{x^*}\nabla^2_{xx}L(w^{\mathrm{a}})\Delta x - \left(2\Delta G^{\mathrm{FE}}(x^*;U^i_{x^*})\bullet\Delta Y^{\mathrm{EF}} + \Delta G^{\mathrm{FF}}(x^*;U^i_{x^*})\bullet\Delta Y^{\mathrm{FF}}\right)_{i=1}^{p_*} = 0.$$

Combined with (A.27) and (A.29), this equation yields

$$U^\top_{x^*}\nabla^2_{xx}L(w^{\mathrm{a}})\Delta x + 2\left(\mathrm{Tr}\left(\Delta G^{\mathrm{FE}}(x^*;U^i_{x^*})Y^{\mathrm{EE}}_{\mathrm{a}}\Delta G^{\mathrm{EF}}(x^*;\Delta x)(G^{\mathrm{FF}}_*)^{-1}\right)\right)_{i=1}^{p_*} = \left(\Delta G^{\mathrm{FF}}(x^*;U^i_{x^*})\bullet(G^{\mathrm{FF}}_*)^{-1}\right)_{i=1}^{p_*}.$$

Decomposing Δx as in (A.26), we obtain from the above equation that

$$\begin{aligned} &U^\top_{x^*}\nabla^2_{xx}L(w^{\mathrm{a}})U_{x^*}\eta^1 + 2\left(\sum_{j=1}^{p_*}\eta^1_j\mathrm{Tr}\left(\Delta G^{\mathrm{FE}}(x^*;U^i_{x^*})Y^{\mathrm{EE}}_{\mathrm{a}}\Delta G^{\mathrm{EF}}(x^*;U^j_{x^*})(G^{\mathrm{FF}}_*)^{-1}\right)\right)_{i=1}^{p_*} \\ &= -U^\top_{x^*}\nabla^2_{xx}L(w^{\mathrm{a}})V_*\eta^2 - 2\left(\mathrm{Tr}\left(\Delta G^{\mathrm{FE}}(x^*;U^i_{x^*})Y^{\mathrm{EE}}_{\mathrm{a}}\Delta G^{\mathrm{EF}}(x^*;V_*\eta^2)(G^{\mathrm{FF}}_*)^{-1}\right)\right)_{i=1}^{p_*} \\ &\quad + \left(-\Delta G^{\mathrm{EF}}(x^*;U^i_{x^*})\bullet(Y^{\mathrm{EE}}_{\mathrm{a}})^{-1} + \left(\Delta G^{\mathrm{EF}}(x^*;U^i_{x^*}) - \Delta G^{\mathrm{FF}}(x^*;U^i_{x^*})\right)\bullet(G^{\mathrm{FF}}_*)^{-1}\right)_{i=1}^{p_*}, \end{aligned} \tag{A.30}$$

where $\eta^1 := (\eta^1_1, \eta^1_2, \ldots, \eta^1_{p_*})^\top$.

Next, we prove that η^1 and η^2 are uniquely determined. To this end, note that $V_*\eta^2 \in \mathcal{U}^\perp_*$ by definition, and $\Delta G^{\mathrm{EE}}(x^*;V_*\eta^2) = (Y^{\mathrm{EE}}_{\mathrm{a}})^{-1}$ follows from (A.28) and $\Delta G^{\mathrm{EE}}(x^*;U_{x^*}\eta^1) = O$. From this, $V_*\eta^2$ turns out to be unique,[6] which together with the full column rank of V_* yields the uniqueness of η^2. In view of this fact and (A.30), η_1 is also uniquely determined, because the matrix

$$U^\top_{x^*}\nabla^2_{xx}L(w^{\mathrm{a}})U_{x^*} + 2\left(\mathrm{Tr}\left(\Delta G^{\mathrm{FE}}(x^*;U^i_{x^*})Y^{\mathrm{EE}}_{\mathrm{a}}\Delta G^{\mathrm{EF}}(x^*;U^j_{x^*})(G^{\mathrm{FF}}_*)^{-1}\right)\right)_{1\le i\le j\le p_*}$$

is actually positive definite by virtue of the SSOSC.

As a result, $\Delta x = U_{x^*}\eta^1 + V_*\eta^2$ is the unique Δx-component of solutions to equations (3.14)-(3.16). Therefore, we ensure Item 1. Item 2 follows immediately from (A.27)-(A.29) with $\Delta x = \xi^*$. □

[6] More precisely speaking, to derive the uniqueness of $V_*\eta^2$, we have made use of the following fundamental result from linear algebra: Given $A \in \mathbb{R}^{q_1\times q_2}$ and $b \in \mathbb{R}^{q_1}$, assume that the linear equation $A\theta = b$ has a nonempty solution set. Pick a solution u arbitrarily and decompose it as $u = u^1 + u^2$ with $u^1 \in \mathrm{Ker}\,A$ and $u^2 \in (\mathrm{Ker}\,A)^\perp$. Then, u^2 is uniquely determined regardless of choice for u, whereas u^1 is free in $\mathrm{Ker}\,A$. In the proof, there exist correspondences between $Au = b$ and $\Delta G^{\mathrm{EE}}(x^*;U_{x^*}\eta^1 + V_*\eta^2) = (Y^{\mathrm{EE}}_{\mathrm{a}})^{-1}$, u^1 and $U_{x^*}\eta^1$, u^2 and $V_*\eta^2$, and $\mathrm{Ker}\,A$ and $\mathcal{U}_*$, respectively.

A.6. Proof of Proposition 6

We prove the first assertion in item 1. Write $\widetilde{X} := P_*^\top X P_*$ for any $X \in \mathbb{S}^m$ as a symbolic rule. In particular, we set $G(\cdot)$ and $\Delta G(x^*;\cdot)$ to X. Denote

$$R(x,\mu) := \widetilde{G}(x) - \widetilde{G}(x^*) - \mu\widetilde{\Delta G}(x^*;\xi^*).$$

We next consider to vary (x,μ) and bound the magnitude of $R(x,\mu)$. Recall $G_*^{\mathrm{EE}} = O$, $G_*^{\mathrm{EF}} = G_*^{\mathrm{FE}} = O$, and $G_*^{\mathrm{FF}} \in \mathbb{S}_{++}^r$. It follows that

$$\begin{aligned}\frac{1}{\mu}\widetilde{G}(x) &= \frac{1}{\mu}\widetilde{G}(x^*) + \widetilde{\Delta G}(x^*;\xi^*) + \frac{1}{\mu}R(x,\mu)\\ &= \begin{bmatrix} (Y_{\mathrm{a}}^{\mathrm{EE}})^{-1} + \frac{1}{\mu}R_1(x,\mu) & \Delta G^{\mathrm{EF}}(x^*;\xi^*) + \frac{1}{\mu}R_2(x,\mu) \\ \Delta G^{\mathrm{FE}}(x^*;\xi^*) + \frac{1}{\mu}R_2(x,\mu)^\top & \Delta G^{\mathrm{FF}}(x^*;\xi^*) + \frac{1}{\mu}G_*^{\mathrm{FF}} + \frac{1}{\mu}R_3(x,\mu)\end{bmatrix},\end{aligned} \tag{A.31}$$

where $R_i(x,\mu)$ $(i = 1,2,3)$ represent block submatrices of $R(x,\mu)$ with appropriate sizes and the second equality follows from $\Delta G^{\mathrm{EE}}(x^*;\xi^*) = (Y_{\mathrm{a}}^{\mathrm{EE}})^{-1}$ by item-2 of Proposition 5. Taylor's expansion of $\widetilde{G}$ at x^* gives

$$\begin{aligned}R(x,\mu) &= \widetilde{\Delta G}\,(x^*, x - x^* - \mu\xi^*) + \mathrm{O}(\|x - x^*\|^2)\\ &= \mathrm{O}(\mu\rho\|\xi^*\| + \|x - x^*\|^2)\\ &= \mathrm{O}(\mu\rho + \mu^2)\end{aligned} \tag{A.32}$$

for $x \in \mathcal{P}_\rho(\mu)$, where the last equality follows since $\|x - x^*\| \le \mu(\rho + 1)\|\xi^*\|$ by $x \in \mathcal{P}_\rho(\mu)$. By (A.32), the fact that $(Y_{\mathrm{a}}^{\mathrm{EE}})^{-1} \in \mathbb{S}_{++}^{m-r_*}$, and taking $\mu_1 > 0$ and $\rho_1 > 0$ so small, $\frac{1}{\mu}R(x,\mu)$ can be so small that the $(1,1)$-block matrix of $\frac{1}{\mu}\widetilde{G}(x)$ is symmetric positive definite for any $(\rho,\mu) \in (0,\rho_1] \times (0,\mu_1]$, that is to say,

$$Q(x,\mu) := (Y_{\mathrm{a}}^{\mathrm{EE}})^{-1} + \frac{1}{\mu}R_1(x,\mu) \in \mathbb{S}_{++}^{m-r_*},\ \forall(\rho,\mu) \in (0,\rho_1] \times (0,\mu_1],\ x \in \mathcal{P}_\rho(\mu), \tag{A.33}$$

which along with (A.32) implies

$$Q(x,\mu)^{-1} = Y_{\mathrm{a}}^{\mathrm{EE}}\,(I + \mathrm{O}\,(\rho + \mu))^{-1}\,. \tag{A.34}$$

Meanwhile, the Schur complement of $\frac{1}{\mu}\widetilde{G}(x)$ is expressed as

$$S_{\mathrm{c}}(x,\mu) := \Delta G^{\mathrm{FF}}(x^*;\xi^*) + \frac{1}{\mu}G_*^{\mathrm{FF}} + \frac{1}{\mu}R_3 - \left(\Delta G^{\mathrm{FE}}(x^*;\xi^*) + \frac{1}{\mu}R_2^\top\right)Q^{-1}\left(\Delta G^{\mathrm{EF}}(x^*;\xi^*) + \frac{1}{\mu}R_2\right),$$

where we have dropped the arguments (x,μ) from the functions R_1, R_2, R_3, and Q for simplicity. From (A.32), by re-taking (μ_1,ρ_1) sufficiently small if necessary, we find that the above S_{c} is symmetric positive definite for any $\mu \in (0,\mu_1]$ because $\frac{1}{\mu}G_*^{\mathrm{FF}} \in S_{++}^{r_*}$ is eventually dominant therein as $\mu > 0$ gets smaller and x moves satisfying $x \in \mathcal{P}_\rho(\mu)$. In particular, the least eigenvalue of $S_{\mathrm{c}}(x,\mu)$ is bounded from below with some positive number, and thus

$$S_{\mathrm{c}}(x,\mu)^{-1} = \mathrm{O}(\mu) \quad (x \in \mathcal{P}_\rho(\mu)). \tag{A.35}$$

As a consequence, $\frac{1}{\mu}\widetilde{G}(x) \in \mathbb{S}_{++}^m$, implying $G(x) \in \mathbb{S}_{++}^m$ for any $(\rho,\mu) \in (0,\rho_1] \times (0,\mu_1]$. Setting $(\bar{\mu},\bar{\rho}) := (\mu_1,\rho_1)$, we ensure the first assertion.

We next prove the second assertion in item 1. Taking the inverse of $\mu^{-1}\widetilde{G}(x)$ by applying the formula of the inverse of a partitioned matrix (e.g., Horn and Johnson [19, Section 0.7.3]) to (A.31), we obtain

$$\mu\widetilde{G}(x)^{-1} = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^\top & M_{22}\end{bmatrix}, \tag{A.36}$$

where each block component is defined as

$$M_{11} := Q(x,\mu)^{-1} + Q(x,\mu)^{-1}\left(\Delta G^{\mathrm{EF}}(x^*;\xi^*) + \frac{1}{\mu}R_2(x,\mu)\right)S_{\mathrm{c}}(x,\mu)^{-1}\left(\Delta G^{\mathrm{EF}}(x^*;\xi^*) + \frac{1}{\mu}R_2(x,\mu)\right)^{\top} Q(x,\mu)^{-1},$$
$$M_{12} := -Q(x,\mu)^{-1}\left(\Delta G^{\mathrm{EF}}(x^*;\xi^*) + \frac{1}{\mu}R_2(x,\mu)\right)S_{\mathrm{c}}(x,\mu)^{-1},$$
$$M_{22} := S_{\mathrm{c}}(x,\mu)^{-1}.$$

Moreover, we have $\frac{1}{\mu}R_i(x,\mu) = \mathrm{O}(\rho+\mu)$ $(i=1,2,3)$ from (A.32). These facts together with (A.34), (A.35), and (A.36) yield

$$M_{11} = Y_{\mathrm{a}}^{\mathrm{EE}}(I + \mathrm{O}(\rho+\mu))^{-1} + \mathrm{O}(\mu),\ M_{12} = \mathrm{O}(\mu),\ M_{22} = \mathrm{O}(\mu), \tag{A.37}$$

which together with $\mu\widetilde{G}(x)^{-1} = \mu P_*^{\top}G(x)^{-1}P_*$ implies that $\{\mu G(x)^{-1} \mid x \in \mathcal{P}_{\bar{\rho}}(\mu), \mu \in (0,\bar{\mu}]\}$ is bounded.

In turn, we will prove item 2. We first consider the first assertion in this item. In what follows, we drive $(x,\mu) \to (x^*,0)$ along with satisfying $x \in \mathcal{P}_{\rho}(\mu)$ and consider limit points and accumulation points of M_{11}, M_{12}, and M_{22}. From (A.37), M_{12} and M_{22} converge to O. Meanwhile, given $\varepsilon > 0$, for any $\rho > 0$ sufficiently small, all accumulation points of M_{11} lie from $Y_{\mathrm{a}}^{\mathrm{EE}}$ within ε. Hence, by recalling the relationship of $\mu\widetilde{G}(x)^{-1}$ and $\mu G(x)^{-1}$, the first assertion is obtained.

Let us turn to the second assertion in item 2. Using (A.34) and the assumption $\rho = o(\mu)$, the convergence of $Q(x,\mu)^{-1}$ to $Y_{\mathrm{a}}^{\mathrm{EE}}$ is ensured. Then, in a similar manner to the first assertion in item 2, the assertion is established with (A.37). The whole proof is complete. □

A.7. Proof of Proposition 7

In this subsection, we prove Proposition 7.

Proof of Proposition 7: In order to prove the first assertion of this proposition, we derive a contradiction by assuming to the contrary, that is, there exists infinite sequences $\{\mu_\ell\} \subseteq \mathbb{R}_{++}$, $\{w^\ell := (x^\ell, Y_\ell, z^\ell)\} \subseteq \bigcup_{\mu\in(0,\mu_\ell]}\mathcal{P}_{\bar{\rho}}(\mu)\times\mathbb{S}^m_{++}\times\mathbb{R}^s$, $\{v^\ell\}\subseteq\mathbb{R}^n$ such that $\lim_{\ell\to\infty}\mu_\ell = 0$ and, for each ℓ, it holds that $\|v^\ell\| = 1$, $\nabla h(x^\ell)^{\top}v^\ell = 0$, (3.19) and (3.20) are fulfilled with $(\mu,w) := (\mu_\ell, w^\ell)$, and

$$\mathcal{H}_\ell := (v^\ell)^{\top}\nabla^2_{xx}L(w^\ell)v^\ell + \Delta G(x^\ell;v^\ell)\bullet\mathcal{L}^{-1}_{G(x^\ell)}\mathcal{L}_{Y_\ell}\left(\Delta G(x^\ell;v^\ell)\right) \le 0. \tag{A.38}$$

From (A.38), it follows that

$$\begin{aligned}
0 &\ge \mathcal{H}_\ell \\
&= (v^\ell)^{\top}\nabla^2_{xx}L(w^\ell)v^\ell + \Delta G(x^\ell;v^\ell)\bullet\mathcal{L}^{-1}_{G_\ell}\mathcal{L}_{\mu_\ell G_\ell^{-1}}\left(\Delta G(x^\ell;v^\ell)\right) + \Delta G(x^\ell;v^\ell)\bullet\mathcal{L}^{-1}_{G_\ell}\mathcal{L}_{Y_\ell-\mu_\ell G_\ell^{-1}}\left(\Delta G(x^\ell;v^\ell)\right) && \text{(A.39)}\\
&= (v^\ell)^{\top}\nabla^2_{xx}L(w^\ell)v^\ell + \mu_\ell\Delta G(x^\ell;v^\ell)\bullet G_\ell^{-1}\Delta G(x^\ell;v^\ell)G_\ell^{-1} + \Delta G(x^\ell;v^\ell)\bullet\mathcal{L}^{-1}_{G_\ell}\mathcal{L}_{Y_\ell-\mu_\ell G_\ell^{-1}}\left(\Delta G(x^\ell;v^\ell)\right) && \text{(A.40)}\\
&= (v^\ell)^{\top}\nabla^2_{xx}L(w^\ell)v^\ell + \mu_\ell\left\|G_\ell^{-\frac{1}{2}}\Delta G(x^\ell;v^\ell)G_\ell^{-\frac{1}{2}}\right\|_{\mathrm{F}}^2 + \Delta G(x^\ell;v^\ell)\bullet\mathcal{L}^{-1}_{G_\ell}\mathcal{L}_{Y_\ell-\mu_\ell G_\ell^{-1}}\left(\Delta G(x^\ell;v^\ell)\right). && \text{(A.41)}
\end{aligned}$$

As will be verified later on, we actually have the following relationships:

$$\lim_{\ell\to\infty}\Delta G(x^\ell;v^\ell)\bullet\mathcal{L}^{-1}_{G_\ell}\mathcal{L}_{Y_\ell-\mu_\ell G_\ell^{-1}}\left(\Delta G(x^\ell;v^\ell)\right) = 0, \tag{A.42}$$
$$\liminf_{\ell\to\infty}\left((v^\ell)^{\top}\nabla^2_{xx}L(w^\ell)v^\ell + \mu_\ell\left\|G_\ell^{-\frac{1}{2}}\Delta G(x^\ell;v^\ell)G_\ell^{-\frac{1}{2}}\right\|_{\mathrm{F}}^2\right) > 0. \tag{A.43}$$

From these results and (A.41), $\liminf_{\ell\to\infty}\mathcal{H}_\ell > 0$ holds. However, this contradicts the hypothesis (A.38). Therefore, we have reached the first assertion.

The second assertion is easily obtained from the first one. Indeed, for proving this, we have only to show that $\mathcal{A}(w)dw = 0$ with $x \in \mathcal{P}_{\rho}(\mu)$ for $dw := (dx, dY, dz)^{\top} \in \mathcal{W}$ implies $dw = 0$. From $\mathcal{A}(w)dw = 0$, it holds that

$$\nabla^2_{xx}L(w)dx - \mathcal{J}G(x)^*dY + \nabla h(x)dz = 0, \tag{A.44}$$
$$\mathcal{L}_{G(x)}dY + \mathcal{L}_Y\Delta G(x;dx) = O, \tag{A.45}$$
$$\nabla h(x)^{\top}dx = 0. \tag{A.46}$$

Note that $(G(x),Y)\in\mathbb{S}^m_{++}\times\mathbb{S}^m_{++}$ and $\nabla h(x)$ is of full column rank under the present setting. Pre-multiplying (A.44) with $dx^\top$ and substituting (A.46) and $dY=-\mathcal{L}^{-1}_{G(x)}\mathcal{L}_Y\Delta G(x;dx)$ from (A.45) into it, we have $dx^\top\nabla^2_{xx}L(w)dx+\Delta G(x;dx)\bullet\mathcal{L}^{-1}_{G(x)}\mathcal{L}_Y(\Delta G(x;dx))=0$. Then, because of Proposition 7, $dx=0$ must hold, which together with (A.45) implies $dY=O$. Moreover, (A.44) and the full column rank of $\nabla h(x)$ give $dz=0$. Hence, we obtain $dw=0$ and thus the second assertion is obtained. □

Proofs of (A.42) **and** (A.43) For making the above proof perfect, it remains to prove (A.42) and (A.43). We will use the same notations and symbols, such as $\{w^\ell\}$, as those in the above proof.

Before going to the proofs of (A.42) and (A.43), we shall give some preliminary results. We first prove that $\{w^\ell\}$ is bounded. The boundedness of $\{x^\ell\}$ is clear because x^ℓ converges to x^* due to $x^\ell\in\bigcup_{\mu\in(0,\mu_\ell]}\mathcal{P}_{\bar\rho}(\mu)$ for each ℓ, and that of $\{Y_\ell\}$ follows from the boundedness of $\{\mu_\ell G_\ell^{-1}\}$ and (3.19). We next prove that $\{z^\ell\}$ is bounded by deriving a contradiction. Suppose to the contrary that there exists a subsequence such that $\|z^\ell\|$ diverges. For simplicity, let us suppose that $\lim_{\ell\to\infty}\|z^\ell\|=\infty$. Condition (3.20) implies $\|\nabla f(x^\ell)-\mathcal{J}G^*Y_\ell+\nabla h(x^\ell)z^\ell\|\le\gamma_2\mu_\ell$ for each ℓ. Dividing both the sides by $\|z^\ell\|$ and driving $\ell\to\infty$, we obtain $\nabla h(x^*)\widetilde{z^*}=0$, where $\widetilde{z^*}$ denotes an accumulation point of $\{z^\ell/\|z^\ell\|\}$ and thus satisfies $\|\widetilde{z^*}\|=1$. Meanwhile, since $\nabla h(x^*)$ is of full column rank because of the MFCQ, $\widetilde{z^*}=0$ must hold. This is a contradiction and hence we conclude $\{z^\ell\}$ is also bounded. Consequently, $\{w^\ell\}$ is bounded.

Let $w^*:=(x^*,Y_*,z^*)$ denote an accumulation point of $\{w^\ell\}$. Notice that w^* is indeed a KKT triplet of the NSDP. Moreover, let v^* be an accumulation point of $\{v^\ell\}$. Choose an orthogonal matrix P_ℓ for each ℓ such that G_ℓ is diagonalized with P_ℓ. Without loss of generality, by re-choosing P_* and taking a subsequence of $\{(x^\ell,v^\ell,P_\ell)\}$ if necessary, we can suppose, w.l.o.g[7],

$$\lim_{\ell\to\infty}(w^\ell,v^\ell,P_\ell)=(w^*,v^*,P_*). \tag{A.47}$$

Note that by definition it follows that

$$\|v^*\|=1,\ \nabla h(x^*)^\top v^*=0. \tag{A.48}$$

Next, so as to match $P_*=[E_*,F_*]$, we partition P_ℓ as

$$P_\ell=[E_\ell,F_\ell],$$

which along with (A.47) implies $\lim_{\ell\to\infty}(E_\ell,F_\ell)=(E_*,F_*)$. Let the resultant diagonal matrix from $G_\ell\in\mathbb{S}^m_{++}$ be D_{G_ℓ}, and also let $D^0_{G_\ell}$ and $D^{++}_{G_\ell}$ be the block diagonal matrices converging to the $(m-r_*)\times(m-r_*)$ zero matrix and the positive diagonal matrix $G^{\rm FF}_*$, respectively. Moreover, we often write $\widetilde{G}_\ell:=P_\ell G_\ell P_\ell^\top$ simply for each ℓ. In summary, it holds that

$$\widetilde{G}_\ell=P_\ell G_\ell P_\ell^\top=\begin{bmatrix}D^0_{G_\ell} & O\\ O & D^{++}_{G_\ell}\end{bmatrix},\ \ \lim_{\ell\to\infty}(D^0_{G_\ell},D^{++}_{G_\ell})=(O,G^{\rm FF}_*).$$

Accordingly, we denote

$$\widetilde{Y}_\ell:=P_\ell Y_\ell P_\ell^\top,\ \widetilde{\Delta G}_\ell:=P_\ell\Delta G_\ell P_\ell^\top$$

with $\Delta G_\ell:=\Delta G(x^\ell;v^\ell)$. Furthermore, so as to match the partition of $\left[\begin{smallmatrix}O & O\\ O & G^{\rm FF}_*\end{smallmatrix}\right]$, partition a given matrix $Z\in\mathbb{S}^m$ as

$$Z=\left[\begin{smallmatrix}Z_{11} & Z_{12}\\ Z_{12}^\top & Z_{22}\end{smallmatrix}\right],\ Z_{11}\in\mathbb{S}^{m-r_*},\ Z_{12}\in\mathbb{R}^{(m-r_*)\times r_*},\ Z_{22}\in\mathbb{S}^{r_*}.$$

Now, we start proving (A.42) and (A.43).

[7] Recall that P_* was an arbitrary orthogonal matrix such that (2.8) holds. Even if P_* is reset as the limit of $\{P_\ell\}$ here, it satisfies (2.8) again, and thus never affects the theoretical results established so far.

Proof of (A.42) First note that, given $W \in \mathbb{S}^m$, a solution $V \in \mathbb{S}^m$ to $\mathcal{L}_{\widetilde{G}_\ell} V = W$ satisfies

$$V_{11} = \mathcal{L}^{-1}_{D^0_{G_\ell}} W_{11},\ V_{22} = \mathcal{L}^{-1}_{D^{++}_{G_\ell}} W_{22}, \tag{A.49}$$

$$V_{12}(i,j) = \frac{1}{D^0_{G_\ell}(i,i) + D^{++}_{G_\ell}(j,j)} W_{12}(i,j) \quad (1 \le i \le m - r_*, 1 \le j \le r_*), \tag{A.50}$$

which are verified by transforming $\mathcal{L}_{\widetilde{G}_\ell} V = W$ as

$$\begin{bmatrix} D^0_{G_\ell} V_{11} + V_{11} D^0_{G_\ell} - W_{11} & D^0_{G_\ell} V_{12} + V_{12} D^{++}_{G_\ell} - W_{12} \\ D^{++}_{G_\ell} V_{12}^\top + V_{12}^\top D^0_{G_\ell} - W_{12}^\top & D^{++}_{G_\ell} V_{22} + V_{22} D^{++}_{G_\ell} - W_{22} \end{bmatrix} = O.$$

By assumption, $Y_\ell - \mu_\ell G_\ell^{-1} = \mathrm{O}(\mu_\ell)$ follows and $\{P_\ell\}$ is bounded, which yields

$$\widetilde{Y}_\ell - \mu_\ell \widetilde{G}_\ell^{-1} = \mathrm{O}(\mu_\ell). \tag{A.51}$$

In view of (A.49) and (A.50) with $W := \mathcal{L}_{\widetilde{Y}_\ell - \mu_\ell \widetilde{G}_\ell^{-1}} \widetilde{\Delta G}_\ell$, we gain that

$$V_{11} = \mathrm{O}(\|\widetilde{\Delta G}_\ell\|_\mathrm{F}),\ V_{22} = \mathrm{O}(\mu_\ell \|\widetilde{\Delta G}_\ell\|_\mathrm{F}),\ V_{12}(i,j) = \mathrm{O}(\mu_\ell \|\widetilde{\Delta G}_\ell\|_\mathrm{F}) \quad (1 \le i \le m - r_*, 1 \le j \le r_*), \tag{A.52}$$

where the first equation in (A.52) is derived from the fact

$$V = \mu_\ell \mathcal{L}^{-1}_{\widetilde{G}_\ell} \mathcal{L}_{\frac{1}{\mu_\ell}(\widetilde{Y}_\ell - \mu_\ell \widetilde{G}_\ell^{-1})} \widetilde{\Delta G}_\ell = \mathrm{O}(\|\widetilde{\Delta G}_\ell\|_\mathrm{F}),$$

which is ensured by (A.51) and the boundedness of $\mu_\ell \mathcal{L}^{-1}_{\widetilde{G}_\ell}$. (For the proof of the boundedness of $\mu_\ell \mathcal{L}^{-1}_{\widetilde{G}_\ell}$, see the footnote [8].) Moreover, the second and third equations in (A.52) are implied by (A.50) and the right equation in (A.49). Using (A.52) again, we obtain

$$\begin{aligned} \Delta G_\ell \bullet \mathcal{L}^{-1}_{G_\ell} \mathcal{L}_{Y_\ell - \mu_\ell G_\ell^{-1}} \Delta G_\ell =& \widetilde{\Delta G}_\ell \bullet \mathcal{L}^{-1}_{\widetilde{G}_\ell} \mathcal{L}_{\widetilde{Y}_\ell - \mu_\ell \widetilde{G}_\ell^{-1}} \widetilde{\Delta G}_\ell \\ =& \widetilde{\Delta G}_\ell \bullet V \\ =& \mathrm{Tr}\left((\widetilde{\Delta G}_\ell)_{11} V_{11} + 2 (\widetilde{\Delta G}_\ell)_{12} V_{12}^\top + (\widetilde{\Delta G}_\ell)_{22} V_{22} \right) \\ =& O\left(\|(\widetilde{\Delta G}_\ell)_{11}\|_\mathrm{F} + \mu_\ell \right). \end{aligned} \tag{A.53}$$

We next prove

$$\Delta G^{\mathrm{EE}}(x^*; v^*) = O. \tag{A.54}$$

As (A.54) deduces $O = \Delta G^{\mathrm{EE}}(x^*; v^*) = \lim_{\ell \to \infty} (\widetilde{\Delta G}_\ell)_{11}$, the desired equation (A.42) follows from (A.53). To prove (A.54), we evaluate $\mu_\ell \left\| G_\ell^{-\frac{1}{2}} \Delta G_\ell G_\ell^{-\frac{1}{2}} \right\|_\mathrm{F}^2$ in (A.41) as follows:

$$\begin{aligned} \mu_\ell \left\| G_\ell^{-\frac{1}{2}} \Delta G_\ell G_\ell^{-\frac{1}{2}} \right\|_\mathrm{F}^2 =& \mu_\ell \mathrm{Tr}\left(G_\ell^{-1} \Delta G_\ell G_\ell^{-1} \Delta G_\ell \right) \\ =& \mu_\ell \mathrm{Tr}\left(\left(\begin{bmatrix} (D^0_{G_\ell})^{-1} & O \\ O & (D^{++}_{G_\ell})^{-1} \end{bmatrix} \begin{bmatrix} (\widetilde{\Delta G}_\ell)_{11} & (\widetilde{\Delta G}_\ell)_{12} \\ (\widetilde{\Delta G}_\ell)_{21} & (\widetilde{\Delta G}_\ell)_{22} \end{bmatrix} \right)^2 \right) \\ =& \mu_\ell \mathrm{Tr}\left(\begin{bmatrix} (D^0_{G_\ell})^{-1} (\widetilde{\Delta G}_\ell)_{11} & (D^0_{G_\ell})^{-1} (\widetilde{\Delta G}_\ell)_{12} \\ (D^{++}_{G_\ell})^{-1} (\widetilde{\Delta G}_\ell)_{21} & (D^{++}_{G_\ell})^{-1} (\widetilde{\Delta G}_\ell)_{22} \end{bmatrix}^2 \right) \\ =& \underbrace{\mu_\ell \mathrm{Tr}\left((D^0_{G_\ell})^{-1} (\widetilde{\Delta G}_\ell)_{11} (D^0_{G_\ell})^{-1} (\widetilde{\Delta G}_\ell)_{11} \right)}_{(a_\ell)} + \underbrace{2 \mu_\ell \mathrm{Tr}\left((D^0_{G_\ell})^{-1} (\widetilde{\Delta G}_\ell)_{12} (D^{++}_{G_\ell})^{-1} (\widetilde{\Delta G}_\ell)_{12}^\top \right)}_{(b_\ell)} \\ &+ \underbrace{\mu_\ell \mathrm{Tr}\left((D^{++}_{G_\ell})^{-1} (\widetilde{\Delta G}_\ell)_{22} (D^{++}_{G_\ell})^{-1} (\widetilde{\Delta G}_\ell)_{22} \right)}_{(c_\ell)}, \end{aligned}$$

[8] Note that, for any $X \in \mathbb{S}^m$ having m eigenvalues $\alpha_1 \le \alpha_2 \le \cdots \le \alpha_m$, the linear operator $\mathcal{L}_X$ has $m(m+1)/2$ eigenvalues $\alpha_1, \alpha_2, \ldots, \alpha_m, \{(\alpha_i + \alpha_j)/2\}_{i \ne j}$. Since $\left\{ \mu G(x)^{-1} \,\middle|\, x \in \mathcal{P}_\rho(\mu), \mu \in (0, \bar{\mu}] \right\}$ is bounded from item 3.5 of Proposition 6, so is $\{\mu_\ell G_\ell^{-1}\}$. Combining these facts, we obtain the desired conclusion immediately.

and herein we obtain

$$\lim_{\ell\to\infty}(c_\ell) = 0 \tag{A.55}$$

because $\lim_{\ell\to\infty}\mu_\ell = 0$ and the matrices in the trace-part of (c_ℓ) are convergent. It holds that $\lim_{\ell\to\infty}\mu_\ell G_\ell^{-1} = Y_*$ since $\mu_\ell G_\ell^{-1}$ and Y_ℓ accumulate at identical points due to (3.19), thus gain $\lim_{\ell\to\infty}\mu_\ell(D^0_{G_\ell})^{-1} = E_*^\top Y_* E_* = (\widetilde{Y}_*)_{11}$. Together with this fact, it follows that

$$\begin{aligned}\lim_{\ell\to\infty}(b_\ell) =& 2\mathrm{Tr}\left((G_*^{\mathrm{FF}})^{-1}\Delta G^{\mathrm{FE}}(x^*;v^*)(\widetilde{Y}_*)_{11}\Delta G^{\mathrm{EF}}(x^*;v^*)\right)\\ =&(v^*)^\top\Omega(x^*,Y_*)v^*,\end{aligned} \tag{A.56}$$

where the last equality follows from Lemma 1. Therefore, in view of (A.41) and (A.53), by noting that $(v^\ell)^\top\nabla^2_{xx}L(w^\ell)v^\ell$ is bounded, we find that $\mu_\ell\left\|G_\ell^{-\frac{1}{2}}\Delta G_\ell G_\ell^{-\frac{1}{2}}\right\|_{\mathrm{F}}^2$ is bounded, which together with (A.55) and (A.55) implies that $\{(a_\ell)\}$ is also bounded. From this fact, we ensure (A.54) since $\mu_\ell(D^0_{G_\ell})^{-1}$ is convergent, and thus the proof of (A.42) is complete.

Proof of (A.43) From (A.54) and (A.48), $v^*\in C(x^*)$ holds, where $C(x^*)$ is the critical cone and expressed as in (2.13). Note that $(Y_*,z^*)\in\Lambda(x^*)$. Hence, by the SSOSC (see (3.1) again for the definition),

$$(v^*)^\top\left(\nabla^2_{xx}L(w^*)+\Omega(x^*,Y_*)\right)v^* > 0. \tag{A.57}$$

Let

$$\Xi_\ell := (v^\ell)^\top\nabla^2_{xx}L(w^\ell)v^\ell + \mu_\ell\left\|G_\ell^{-\frac{1}{2}}\Delta G(x^\ell;v^\ell)G_\ell^{-\frac{1}{2}}\right\|_{\mathrm{F}}^2$$

for each ℓ. Since $(a_\ell)\geq 0$ for each ℓ and $\{(a_\ell)\}$ is bounded as shown in the proof of (A.42), any accumulation point of $\{(a_\ell)\}$ is nonnegative. Combining this fact together with (A.55), (A.56), and (A.57) yields

$$\liminf_{\ell\to\infty}\Xi_\ell = (v^*)^\top\left(\nabla^2_{xx}L(w^*)+\Omega(x^*,Y_*)\right)v^* + \liminf_{\ell\to\infty}(a_\ell) > 0,$$

which implies (A.43). The proof is complete.

□

A.8. Proof of Lemma 2

In this subsection, we prove Lemma 2.

Proof of Lemma 2: Recall that $\nabla h(x^*)$ is of full column rank. First, we show that

$$\mathrm{dist}\,(x,\mathcal{M}) = \mathrm{O}(\|h(x)\|) \tag{A.58}$$

for any $x\in\mathbb{R}^n$ locally around x^*. Let $C\subseteq\mathbb{R}^n$ be a closed ball centered at x^*, such that $\nabla h(x)$ is of full column rank for all $x\in C$, namely,

$$\mathrm{rank}\nabla h(x) = \mathrm{rank}\nabla h(x^*) = s,\ \forall x\in C. \tag{A.59}$$

To prove (A.58), assume to the contrary: There exists a convergent sequence $\{x^l\}\subseteq C$ converging to some $\bar{x}\in C$ such that $\|h(x^l)\|/\mathrm{dist}\,(x^l,\mathcal{M})\to 0$ as $l\to\infty$. For each l, let $y^l\in\arg\min_{y\in\mathcal{M}\cap C}\|x^l-y\|$. We then

$$\mathrm{dist}\,(x^l,\mathcal{M}) = \|x^l-y^l\|$$

for any l large enough. Notice $y^l\to\bar{x}$ and thus $x^l-y^l\to 0$. Since y^l solves $\min_{y\in\mathcal{M}\cap C}\|x^l-y\|^2$ by definition and y^l stays in the interior of C for any l large enough, by the KKT conditions there exists η^l for each l large enough such that

$$x^l-y^l = \nabla h(y^l)\eta^l.$$

Letting $\bar{v}$ be an accumulation point of $\{(x^l - y^l)/\|x^l - y^l\|\}$, we may assume that

$$\frac{(x^l - y^l)}{\|x^l - y^l\|} = \frac{\nabla h(y^l)\eta^l}{\|\nabla h(y^l)\eta^l\|} \to \bar{v} \ \ (l \to \infty) \tag{A.60}$$

without loss of generality. It follows that $\|\bar{v}\| = 1$. As

$$\begin{aligned}\frac{\|h(x^l)\|}{\operatorname{dist}(x^l, \mathcal{M})} &= \frac{\|h(x^l) - h(y^l)\|}{\|x^l - y^l\|} \\ &= \frac{\|\nabla h(y^l)^\top (x^l - y^l) + \mathrm{O}(\|x^l - y^l\|^2)\|}{\|x^l - y^l\|}\end{aligned}$$

for any l large enough and by driving $l \to \infty$ here, we obtain

$$\nabla h(\bar{x})^\top \bar{v} = 0. \tag{A.61}$$

Note $y^l \to \bar{x}$ holds. Since $\nabla h(\bar{x})$ is of full column rank because of $\bar{x} \in C$ and (A.59), (A.61) actually implies that there exists $\{v^l\}$ such that it converges to $\bar{v}$ and, for each l,

$$\nabla h(y^l)^\top v^l = 0. \tag{A.62}$$

Meanwhile, with (A.60) and $\lim_{l\to\infty} v^l = \bar{v}$, we gain

$$\lim_{l\to\infty} \frac{(v^l)^\top \nabla h(y^l)\eta^l}{\|\nabla h(y^l)\eta^l\|} = \|\bar{v}\|^2,$$

which combined with (A.62) gives $\|\bar{v}\|^2 = 0$. This however contradicts $\|\bar{v}\| = 1$. We thus obtain (A.58).

We next prove the desired relation (3.24). Note $h(x^* + \mu\xi^*) = \mathrm{O}(\mu^2)$ holds from Taylor's expansion together with $h(x^*) = 0 = \nabla h(x^*)^\top \xi^*$. With this fact along with (A.58), (3.24) is ensured. The proof is complete. □

References

[1] Andreani R, Fukuda E, Haeser G, Santos D, Secchin L (2021) On the use of Jordan Algebras for improving global convergence of an augmented Lagrangian method in nonlinear semidefinite programming. *Computational Optimization and Applications* 1–16.

[2] Andreani R, Haeser G, Viana DS (2018) Optimality conditions and global convergence for nonlinear semidefinite programming. *Mathematical Programming* 1–33.

[3] Arahata S, Okuno T, Takeda A (2021) Interior-point methods for second-order stationary points of nonlinear semidefinite optimization problems using negative curvature. *arXiv preprint arXiv:2103.14320* .

[4] Auslender A (2013) An extended sequential quadratically constrained quadratic programming algorithm for nonlinear, semidefinite, and second-order cone programming. *Journal of Optimization Theory and Applications* 156(2):183–212.

[5] Auslender A (2015) An exact penalty method for nonconvex problems covering, in particular, nonlinear programming, semidefinite programming, and second-order cone programming. *SIAM Journal on Optimization* 25(3):1732–1759.

[6] Bonnans JF, Ramírez H (2005) Perturbation analysis of second-order cone programming problems. *Mathematical Programming* 104(2):205–227.

[7] Bonnans JF, Shapiro A (2013) *Perturbation analysis of optimization problems* (Springer Science & Business Media).

[8] Correa R, Ramirez C H (2004) A global algorithm for nonlinear semidefinite programming. *SIAM Journal on Optimization* 15(1):303–318.

[9] da Cruz Neto JX, Ferreira OP, Monteiro RD (2005) Asymptotic behavior of the central path for a special class of degenerate SDP problems. *Mathematical Programming* 103(3):487–514.

[10] Forsgren A (2000) Optimality conditions for nonconvex semidefinite programming. *Mathematical Programming* 88(1):105–128.

[11] Freund RW, Jarre F, Vogelbusch CH (2007) Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling. *Mathematical Programming* 109(2-3):581–611.

[12] Fukuda EH, Lourenço BF (2018) Exact augmented Lagrangian functions for nonlinear semidefinite programming. *Computational Optimization and Applications* 71(2):457–482.

[13] Goldfarb D, Scheinberg K (1998) Interior point trajectories in semidefinite programming. *SIAM Journal on Optimization* 8(4):871–886.

[14] Graña Drummond L, Peterzil Y (2002) The central path in smooth convex semidefinite programs. *Optimization* 51(2):207–233.

[15] Halická M (2002) Analyticity of the central path at the boundary point in semidefinite programming. *European Journal of Operational Research* 143(2):311–324.

[16] Halická M, de Klerk E, Roos C (2002) On the convergence of the central path in semidefinite optimization. *SIAM Journal on Optimization* 12(4):1090–1099.

[17] Halická M, de Klerk E, Roos C (2005) Limiting behavior of the central path in semidefinite optimization. *Optimization Methods and Software* 20(1):99–113.

[18] Hoi C, Scherer CW, Van der Meché E, Bosgra O (2003) A nonlinear SDP approach to fixed-order controller synthesis and comparison with two other methods applied to an active suspension system. *European Journal of Control* 9(1):13–28.

[19] Horn RA, Johnson CR (2012) *Matrix Analysis* (Cambridge University Press).

[20] Huang X, Teo K, Yang X (2006) Approximate augmented Lagrangian functions and nonlinear semidefinite programs. *Acta Mathematica Sinica* 22(5):1283–1296.

[21] Jarre F (2000) An interior method for nonconvex semidefinite programs. *Optimization and Engineering* 1(4):347–372.

[22] Kakihara S, Ohara A, Tsuchiya T (2013) Information geometry and interior-point algorithms in semidefinite programs and symmetric cone programs. *Journal of Optimization Theory and Applications* 157(3):749–780.

[23] Kakihara S, Ohara A, Tsuchiya T (2014) Curvature integrals and iteration complexities in SDP and symmetric cone programs. *Computational Optimization and Applications* 57(3):623–665.

[24] Kanzow C, Nagel C, Kato H, Fukushima M (2005) Successive linearization methods for nonlinear semidefinite programs. *Computational Optimization and Applications* 31(3):251–273.

[25] Kato A, Yabe H, Yamashita H (2015) An interior point method with a primal–dual quadratic barrier penalty function for nonlinear semidefinite programming. *Journal of Computational and Applied Mathematics* 275:148–161.

[26] Kato H, Fukushima M (2007) An SQP-type algorithm for nonlinear second-order cone programs. *Optimization Letters* 1(2):129–144.

[27] Kočvara M, Leibfritz F, Stingl M, Henrion D (2005) A nonlinear SDP algorithm for static output feedback problems in COMPleib. *IFAC Proceedings Volumes* 38(1):1055–1060.

[28] Kočvara M, Stingl M (2004) Solving nonconvex SDP problems of structural optimization with stability control. *Optimization Methods and Software* 19(5):595–609.

[29] Kojima M, Mizuno S, Noma T (1990) Limiting behavior of trajectories generated by a continuation method for monotone complementarity problems. *Mathematics of Operations Research* 15(4):662–675.

[30] Konno H, Kawadai N, Tuy H (2003) Cutting plane algorithms for nonlinear semi-definite programming problems with applications. *Journal of Global Optimization* 25(2):141–155.

[31] Leibfritz F, Maruhn JH (2009) A successive SDP-NSDP approach to a robust optimization problem in finance. *Computational Optimization and Applications* 44(3):443.

[32] Leibfritz F, Mostafa E (2002) An interior point constrained trust region method for a special class of nonlinear semidefinite programming problems. *SIAM Journal on Optimization* 12(4):1048–1074.

[33] Leibfritz F, Volkwein S (2006) Reduced order output feedback control design for PDE systems using proper orthogonal decomposition and nonlinear semidefinite programming. *Linear Algebra and its Applications* 415(2-3):542–575.

[34] Lourenço BF, Fukuda EH, Fukushima M (2018) Optimality conditions for nonlinear semidefinite programming via squared slack variables. *Mathematical Programming* 168(1-2):177–200.

[35] Luenberger DG, Ye Y (2008) *Linear and Nonlinear Programming, Third Edition* (Springer Science & Business Media).

[36] Luo ZQ, Sturm JF, Zhang S (1998) Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming. *SIAM Journal on Optimization* 8(1):59–81.

[37] Mangasarian OL (1994) *Nonlinear programming* (SIAM).

[38] Megiddo N (1989) Pathways to the optimal set in linear programming. *Progress in mathematical programming*, 131–158 (Springer).

[39] Monteiro RD, Tsuchiya T (1996) Limiting behavior of the derivatives of certain trajectories associated with a monotone horizontal linear complementarity problem. *Mathematics of Operations Research* 21(4):793–814.

[40] Monteiro RD, Zou F (1998) On the existence and convergence of the central path for convex programming and some duality results. *Computational Optimization and Applications* 10(1):51–77.

[41] Okuno T (2020) Local convergence of primal-dual interior point methods for nonlinear semi-definite optimization using the family of Monteiro-Tsuchiya directions. *arXiv preprint arXiv:2009.03020* .

[42] Okuno T, Fukushima M (2020) An interior point sequential quadratic programming-type method for log-determinant semi-infinite programs. *Journal of Computational and Applied Mathematics* 376:112784.

[43] Okuno T, Fukushima M (2022) Primal-dual path following method for nonlinear semi-infinite programs with semi-definite constraints. *Mathematical Programming* 1–53.

[44] Qi H (2009) Local duality of nonlinear semidefinite programming. *Mathematics of Operations Research* 34(1):124–141.

[45] Qi H, Sun D (2006) A quadratically convergent Newton method for computing the nearest correlation matrix. *SIAM Journal on Matrix Analysis and Applications* 28(2):360–385.

[46] Scherer CW (1995) Multiobjective H_2/H_∞control. *IEEE Transactions on Automatic Control* 40(6):1054–1062.

[47] Shapiro A (1997) First and second order analysis of nonlinear semidefinite programs. *Mathematical Programming* 77(1):301–320.

[48] Sturm JF, Zhang S (2001) On sensitivity of central solutions in semidefinite programming. *Mathematical Programming* 90(2):205–227.

[49] Sun D (2006) The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Mathematics of Operations Research* 31(4):761–776.

[50] Sun D, Sun J, Zhang LW (2008) The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. *Mathematical Programming* 114(2):349–391.

[51] Sun J, Zhang LW, Wu Y (2006) Properties of the augmented Lagrangian in nonlinear semidefinite optimization. *Journal of Optimization Theory and Applications* 129(3):437–456.

[52] Takezawa A, Nii S, Kitamura M, Kogiso N (2011) Topology optimization for worst load conditions based on the eigenvalue analysis of an aggregated linear system. *Computer Methods in Applied Mechanics and Engineering* 200(25-28):2268–2281.

[53] Thore CJ (2022) A worst-case approach to topology optimization for maximum stiffness under uncertain boundary displacement. *Computers & Structures* 259:106696.

[54] Thore CJ, Holmberg E, Klarbring A (2017) A general framework for robust topology optimization under load-uncertainty including stress constraints. *Computer Methods in Applied Mechanics and Engineering* 319:1–18.

[55] Vandaele A, Glineur F, Gillis N (2018) Algorithms for positive semidefinite factorization. *Computational Optimization and Applications* 71(1):193–219.

[56] Vandenberghe L, Boyd S (1996) Semidefinite programming. *SIAM Review* 38(1):49–95.

[57] Wolkowicz H, Saigal R, Vandenberghe L (2012) *Handbook of semidefinite programming: theory, algorithms, and applications*, volume 27 (Springer Science & Business Media).

[58] Wright SJ, Orban D (2002) Properties of the log-barrier function on degenerate nonlinear programs. *Mathematics of Operations Research* 27(3):585–613.

[59] Wu H, Luo H, Ding X, Chen G (2013) Global convergence of modified augmented Lagrangian methods for nonlinear semidefinite programming. *Computational Optimization and Applications* 56(3):531–558.

[60] Yamakawa Y, Okuno T (2022) A stabilized sequential quadratic semidefinite programming method for degenerate nonlinear semidefinite programs. *Computational Optimization and Applications* 1–38.

[61] Yamakawa Y, Yamashita N (2014) A two-step primal-dual interior point method for nonlinear semidefinite programming problems and its superlinear convergence. *Journal of the Operations Research Society of Japan* 57(3-4):105–127.

[62] Yamakawa Y, Yamashita N (2015) A differentiable merit function for the shifted perturbed KKT conditions of the nonlinear semidefinite programming. *Pacific Journal of Optimization* 11(3):557–579.

[63] Yamashita H, Yabe H (2012) Local and superlinear convergence of a primal-dual interior point method for nonlinear semidefinite programming. *Mathematical Programming* 132(1-2):1–30.

[64] Yamashita H, Yabe H, Harada K (2012) A primal–dual interior point method for nonlinear semidefinite programming. *Mathematical Programming* 135(1-2):89–121.

[65] Yamashita H, Yabe H, Harada K (2021) A primal-dual interior point trust-region method for nonlinear semidefinite programming. *Optimization Methods and Software* 36(2-3):569–601.

[66] Yang L, Yu B (2013) A homotopy method for nonlinear semidefinite programming. *Computational Optimization and Applications* 56(1):81–96.

[67] Zhao Q, Chen Z (2016) On the superlinear local convergence of a penalty-free method for nonlinear semidefinite programming. *Journal of Computational and Applied Mathematics* 308:1–19.

[68] Zhao Q, Chen Z (2018) An SQP-type method with superlinear convergence for nonlinear semidefinite programming. *Asia-Pacific Journal of Operational Research* 35(03):1850009.