Noname manuscript No.
(will be inserted by the editor)

Stratification for Nonlinear Semidefinite Programming

Chenglong Bao · Chao Ding · Fuxiaoyue Feng · Jingyu Li

This version: January 13, 2026

Abstract This paper introduces a stratification framework for nonlinear semidefinite programming (NLSDP) that reveals and utilizes the geometry behind the nonsmooth KKT system. Based on the *index stratification* of $\mathbb{S}^n$ and its lift to the primal–dual space, a stratified variational analysis is developed. Specifically, we define the stratum-restricted regularity property, characterize it by the verifiable weak second order condition (W-SOC) and weak strict Robinson constraint qualification (W-SRCQ), and interpret the W-SRCQ geometrically via transversality, which provides its genericity along ambient space and stability along strata. The interactions of these properties across neighboring strata are further examined, leading to the conclusion that classical strong-form regularity conditions correspond to the local uniform validity of stratum-restricted counterparts. On the algorithmic side, a stratified Gauss–Newton method with normal steps and a correction mechanism is proposed for globally solving the KKT equation through a least-squares merit function. We demonstrate that the algorithm converges globally to directional stationary points. Moreover, under the W-SOC and the strict Robinson constraint qualification (SRCQ), it achieves local quadratic convergence to KKT pairs and eventually identifies the active stratum.

Key words: nonlinear semidefinite programming; stratification; variational analysis; perturbation analysis; strong metric regularity; Gauss–Newton.

1 Introduction

Consider the nonlinear semidefinite programming (NLSDP) problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \in \mathbb{S}^n_+, \end{aligned} \tag{1}$$

where $\mathbb{X}$ is a Euclidean space, $f : \mathbb{X} \to \mathbb{R}$ and $g : \mathbb{X} \to \mathbb{S}^n$ are twice continuously differentiable, and $\mathbb{S}^n_+$ denotes the cone of symmetric positive semidefinite matrices. The NLSDP (1) is a fundamental optimization

C.L. Bao
Yau Mathematical Sciences Center, Tsinghua University, Beijing 100084, P.R. China; Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, P.R. China.
E-mail: clbao@tsinghua.edu.cn

C. Ding
State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P.R. China; School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, P.R. China; Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P.R. China
E-mail: dingchao@amss.ac.cn

F.X.Y. Feng
Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China, School of Mathematical Sciences, University of Chinese Academy of Science, Beijing 100049, P.R. China.
E-mail: fengfuxiaoyue@amss.ac.cn

J.Y. Li
Qiuzhen College, Tsinghua University, Beijing 100084, P.R. China.
E-mail: lijingyu23@mails.tsinghua.edu.cn

model that includes nonlinear programming and semidefinite programming as special cases. It arises in a wide range of applications, including control and systems theory, structural optimization, material design, eigenvalue optimization, and robust optimization, and has motivated extensive work in variational analysis, perturbation theory, and numerical optimization. See, e.g., [6] and the references therein.

A pair $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$ is a KKT pair of (1) if and only if it solves a generally nonsmooth equation of the form

$$F(z) := \begin{bmatrix} \nabla f(x) + \nabla g(x)\, y \\ -g(x) + \Pi_{\mathbb{S}^n_+}\big(G(z)\big) \end{bmatrix} = 0, \tag{2}$$

where $\Pi_{\mathbb{S}^n_+} : \mathbb{S}^n \to \mathbb{S}^n$ is the metric projection onto $\mathbb{S}^n_+$ and $G(z) := g(x) + y$. The nonsmoothness in (2) comes from $\Pi_{\mathbb{S}^n_+}$, which is Lipschitz and directionally differentiable but not differentiable at matrices with zero eigenvalues. As a consequence, even when f and g are smooth, the KKT mapping F is nonsmooth, and classical smooth Newton theory does not apply directly.

Semismooth Newton-type methods [34,33] provide a natural tool for solving (2). Their fast local convergence is typically ensured under the *strong regularity* [35, Section 2]. In particular, for the NLSDP (1), Sun [39] shows that under the Robinson constraint qualification (RCQ, Definition 1), for a local optimal solution, the strong regularity is known to be equivalent to a combination of the constraint nondegeneracy (Definition 3) and strong second order sufficient condition (S-SOSC, Definition 5). However, these strong-form conditions may fail in applications when degeneracy appears. More recently, Feng et al. [19] introduced the weak strict Robinson constraint qualification (W-SRCQ, Definition 6) and the weak second order condition (W-SOC, Definition 7), and proposed a correction mechanism that yields superlinear or quadratic local behavior. However, these weak conditions are originally introduced from an algorithmic perspective, and their relationship to the classical perturbation and regularity theory for the NLSDP has not been fully clarified. Moreover, a geometric interpretation that explains why such conditions are natural, and how they behave under perturbations, remains largely implicit. The present work shall fill this gap.

In recent years, solving (1) under the assumption that the solutions are low-rank, commonly referred to as low-rank SDP, has also attracted significant attention. Manifold optimization, which treats the rank constraint from a geometric perspective, has emerged as a powerful approach. In this framework, the search is restricted to the fixed-rank manifold $\mathcal{M}_r := \{X \in \mathbb{S}^n_+ : \mathrm{rank(X)} = \mathrm{r}\}$, which forms a smooth Riemannian submanifold of the space of symmetric matrices. Optimization over $\mathcal{M}_r$ enables the use of intrinsic algorithms such as the Riemannian Newton method, which achieves local quadratic convergence provided the Riemannian Hessian is nondegenerate at the solution and a suitable retraction is employed [1]. However, this approach requires prior knowledge of the true rank r^*. Otherwise, it encounters the manifold identification problem, as discussed in [46, 20]. As a complementary strategy, the Burer–Monteiro (BM) factorization $X = RR^\top$ with $R \in \mathbb{R}^{n\times r}$ recasts the low-rank SDP as a nonconvex optimization problem in Euclidean space [9,10]. If r is chosen such that $r(r+1)/2 > m$, where m denotes the number of affine constraints, the BM formulation admits no spurious local minima [8]. Nevertheless, choosing r excessively large leads to ill-conditioning and higher computational cost, illustrating a fundamental trade-off in the BM approach between theoretical exactness and practical efficiency.

In this paper, our focus is not on optimizing a smooth function over a prescribed manifold. Instead, we analyze a nonsmooth problem by decomposing the ambient space into smooth manifolds on which the relevant mappings are smooth. A related perspective appears in the theory of *partial smoothness* [28], where the Clarke regularity and normal sharpness provide a well-controlled local geometry along a distinguished active manifold. However, the functions considered here do not satisfy the normal sharpness condition, due to the more intricate interplay of nonconvexity and nondifferentiability in their local geometry. This calls for a finer geometric description, which is provided here through a *stratification* framework for (1).

Stratification is a mathematical technique for decomposing a singular space (e.g., a nonsmooth set or variety) into a collection of smooth manifolds called strata. Originally developed to study singularities of mappings and spaces, it has become a fundamental tool in many areas, including algebraic geometry [43,44,21], singularity theory [2,30], and notably, nonsmooth optimization [16,24]. The formal concept of stratification emerged in the mid-20th century from the work of Hassler Whitney, René Thom and John Mather, who sought to understand the nature of singular spaces. Whitney [43,44] introduced geometric regularity conditions that ensure the stratified space has a locally trivial structure along each stratum, a property crucial for detailed analysis. Thom [41] proved that these conditions are precisely what you need to ensure the space is topologically "rigid" enough to study systematically. The theory was significantly extended in the 1970s by Mather (see [30]), who provided a detailed account of control data and the stability of stratified mappings. The framework was also generalized to broader classes of sets, such as definable sets in o-minimal structures, thereby making it applicable to a wide range of functions and constraints encountered in applied mathematics. A significant application of stratification is in nonsmooth optimization, particularly within the framework of tame optimization. Along this line of thought

lies a rich set of interesting works, which is far beyond the scope of this paper. Interested readers may refer to [24] for further details.

In this paper, our use of stratification is not a modeling assumption on the feasible set. Instead, it is an analytic lens for the KKT system: once the spectral pattern of a symmetric matrix is fixed, the projector $\Pi_{\mathbb{S}^n_+}$ becomes smooth. This leads to the *index stratification* of $\mathbb{S}^n$ into manifolds of matrices with fixed numbers of positive and negative eigenvalues, and, via the mapping $G(z) = g(x) + y$, to a lifted stratification of $\mathbb{X} \times \mathbb{S}^n$ on which the KKT mapping is smooth on each stratum and admits an explicit differential. This observation offers a new possibility for designing a global convergent Newton method with a quadratic local convergence rate under much weaker conditions. Moreover, the stratification framework of (1) allows us to revisit classical variational and perturbation theory from a stratified perspective, which surprisingly reveals some fundamental properties of classical perturbation results of (1).

Perturbation analysis is a cornerstone of modern optimization and variational analysis: it explains how stationary points and multipliers respond to small changes in data, and it underpins error bounds and convergence-rate guarantees for numerical algorithms. A central object in this theory is the KKT solution mapping and its regularity properties, such as strong regularity, which provide quantitative stability estimates of Lipschitz type; see, e.g., the unified treatment in [15,6]. In nonlinear programming, the strong regularity is characterized by the S-SOSC together with the LICQ [35,25,7,14], while in nonpolyhedral conic optimization (e.g., the NLSDP (1)), analogous characterizations have been established. In particular, under the RCQ, strong regularity of (1) is known [39] to be equivalent to the S-SOSC together with constraint nondegeneracy at a local optimum, as stated previously. Such results are fundamental, but they typically treat the KKT system as a single nonsmooth object and therefore lead naturally to conditions ensuring nonsingularity of generalized Jacobians in a neighborhood or on error bound assumptions that are difficult to verify in practice and may be overly conservative for algorithmic purposes. This tension is particularly visible in the analysis of algorithms whose rate statements hinge on error bounds or calmness-type properties of KKT solution mappings, such as the augmented Lagrange method and Newton-type schemes.

Within the stratification framework, our analysis proceeds on two intertwined levels. On a fixed stratum, the stratum-wise smoothness of $\Pi_{\mathbb{S}^n_+}$ allows us to treat the KKT mapping as a smooth equation on a smooth manifold and to compute its manifold differential in closed form. This, in turn, motivates a *stratum-restricted* notion of strong metric regularity, tailored to the geometry induced by the index stratification. We show that, on each fixed stratum, this stratum-restricted regularity admits simple and verifiable characterizations and, in particular, is governed precisely by the weak conditions introduced in [19]. Moreover, by interpreting these conditions through transversality, we obtain a geometric explanation of their stability and genericity along strata. Beyond the stratum-wise theory, we study how regularity properties propagate across adjacent strata. Here, the stratified viewpoint offers a refined perturbation picture: the weak-form conditions are stable under perturbations that preserve the relevant inertia indices, while the classical strong-form conditions emerge as the uniform validity over a neighborhood of the weak-form conditions. It is worth mentioning that this stratification-based perspective is not tied to Newton-type methods. Since stability of KKT solution mappings also drives the local theory of augmented Lagrangian and other primal–dual schemes, the stratum-restricted perturbation viewpoint developed here suggests a route to relax and sharpen assumptions in the analysis of a broader class of algorithms for (1) and related nonpolyhedral problems.

On the algorithmic side, we use the stratification framework to design a global method *Stratified Gauss–Newton with correction* (SGN) method for solving the KKT equation (2) via the nonsmooth least-squares merit function $\varphi(z) = \frac{1}{2}\|F(z)\|^2$. The resulting scheme combines (i) a Levenberg–Marquardt step along the current stratum, (ii) carefully chosen normal-direction moves that enable escape from stationary behavior confined to a stratum, and (iii) an eigenvalue-triggered correction mechanism for identifying a more appropriate spectral pattern. A merit-function-based acceptance rule guarantees monotone decrease of φ and global convergence to a directional stationary (D-stationary) point (Definition 20), and the local analysis exploits the stratified perturbation theory to show that the iterates identify the active stratum and converge quadratically under the W-SOC and the SRCQ.

The remainder of the paper is organized as follows. Section 2 summarizes background material on the projector onto $\mathbb{S}^n_+$, standard regularity conditions for NLSDP, and the basic notions from differential geometry and stratification used throughout. In Section 3, we first introduce the index stratification of $\mathbb{S}^n$, prove that $\Pi_{\mathbb{S}^n_+}$ is C^∞ when restricted to each stratum, and derive explicit differential formulas; we further examine the stratum-wise variational theory, both within a fixed stratum and across strata, including stratum-restricted regularity and its characterizations via the W-SOC and W-SRCQ, as well as the related stability, genericity, and transversality interpretations. Section 4 proposes a global stratified Gauss–Newton scheme and establishes global convergence together with local quadratic convergence. We conclude this paper in Section 5. Technical proofs are provided in the appendices.

2 Preliminaries

To begin with, we introduce the notation system of this paper. $\mathbb{X} := \mathbb{R}^m$ denotes a Euclidean space of dimension m; $\mathbb{S}^n$ denotes the space of $n\times n$ symmetric matrices; $\mathbb{S}^n_+$ denotes the cone of $n\times n$ symmetric positive semidefinite matrices; and $\Pi_{\mathbb{S}^n_+}$ denotes the projection operator onto $\mathbb{S}^n_+$. For a smooth mapping F between Euclidean spaces, $F'(x)$ denotes its Fréchet derivative (Jacobian), and $\nabla F(x)$ denotes its gradient. For a directionally differentiable function f on a Euclidean space, $f'(x;d)$ denotes the directional derivative of f at x in the direction d.

Other notations are listed below.

- A_{ij}: the (i,j)-th entry of a matrix $A \in \mathbb{R}^{m\times n}$.
- A_i: the i-th column of a matrix $A \in \mathbb{R}^{m\times n}$.
- $A_{\alpha\beta}$: the $|\alpha| \times |\beta|$ submatrix of A obtained by retaining only the rows indexed by the index set α, and the columns indexed by the index set β.
- "$\circ$": Hadamard product between matrices, i.e., for any two matrices A and B in $\mathbb{R}^{m\times n}$ the (i,j)-th entry of $Z := A \circ B \in \mathbb{R}^{m\times n}$ is $Z_{ij} = A_{ij}B_{ij}$.
- $\mathrm{Diag}(\cdot)$: the operator that maps a vector to a diagonal matrix with the vector as its diagonal entries.
- $\mathrm{diag}(\cdot)$: the operator that maps a matrix to a vector consisting of its diagonal entries.
- $\mathrm{lin}(C)$: the linearity space of a closed convex set $C \subseteq \mathbb{X}$, i.e., the largest subspace in C.
- $\mathrm{aff}(C)$: the affine hull of C, i.e., the smallest subspace containing C.
- $\mathcal{M}_{p,q}$ (or $\mathcal{M}$): the set of symmetric matrices with exactly p positive and q negative eigenvalues.
- $\widetilde{\mathcal{M}}_{p,q}$ (or $\widetilde{\mathcal{M}}$): the preimage $G^{-1}(\mathcal{M}_{p,q})$ (see (21)).

2.1 Variational properties of $\Pi_{\mathbb{S}^n_+}$

In this section, we collect some basic properties of the metric projection $\Pi_{\mathbb{S}^n_+}$. Let $A \in \mathbb{S}^n$ be given. Denote the eigenvalues of A (all real and counted with multiplicities) in nonincreasing order by

$$\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A),$$

and let $\lambda(A)$ be the vector of these ordered eigenvalues. Define $\Lambda(A) := \mathrm{Diag}(\lambda(A))$. Let $\alpha(A)$, $\beta(A)$ and $\gamma(A)$ be the index sets defined by

$$\alpha(A) := \{i \mid \lambda_i(A) > 0\}, \quad \beta(A) := \{i \mid \lambda_i(A) = 0\}, \quad \gamma(A) := \{i \mid \lambda_i(A) < 0\}. \tag{3}$$

In later discussions, when the dependence of α, β, and γ on A can be seen clearly from the context, we often drop A from these notations. Denote the positive and negative inertia indices of A by $p := |\alpha(A)|$ and $q := |\gamma(A)|$, respectively. Consider the eigenvalue decomposition of A

$$A = \begin{bmatrix} P_\alpha & P_\beta & P_\gamma \end{bmatrix} \begin{bmatrix} \Lambda(A)_{\alpha\alpha} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda(A)_{\gamma\gamma} \end{bmatrix} \begin{bmatrix} P_\alpha^\top \\ P_\beta^\top \\ P_\gamma^\top \end{bmatrix}, \tag{4}$$

where $P \in \mathcal{O}^n$. Let $\mathcal{O}^n(A)$ be the set of all orthogonal matrices P satisfying (4). We refer to the collection $(\alpha, \beta, \gamma, p, q, P, \lambda)$ as an *indexed eigenvalue decomposition (IED)* of A.

Remark 1 We say "an" IED of A, rather than "the" IED of A, since the matrix $P \in \mathcal{O}^n(A)$ is generally not unique. This nonuniqueness will be taken into account throughout the paper. In particular, most of our subsequent constructions and results are invariant with respect to the choice of P.

It is well-known that the metric projection $\Pi_{\mathbb{S}^n_+}(A) = P_\alpha\, \Lambda(A)_{\alpha\alpha}\, P_\alpha^\top$. Moreover, $\Pi_{\mathbb{S}^n_+}$ is B-differentiable (that is, locally Lipschitz and directionally differentiable; see [4,5]) and semismooth [40], but in general not (Fréchet) differentiable. More precisely, the directional derivative of $\Pi_{\mathbb{S}^n_+}$ at A in the direction $H \in \mathbb{S}^n$ is given by

$$\Pi'_{\mathbb{S}^n_+}(A;H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \Xi_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\alpha\beta}^\top & \Pi_{\mathbb{S}^{|\beta|}_+}(\widetilde{H}_{\beta\beta}) & 0 \\ \Xi_{\alpha\gamma}^\top \circ \widetilde{H}_{\alpha\gamma}^\top & 0 & 0 \end{bmatrix} P^\top, \tag{5}$$

where $P \in \mathcal{O}^n(A)$, $\widetilde{H} := P^\top H P$, and $\Xi \in \mathbb{S}^n$ is defined by

$$\Xi_{ij} := \frac{\max\{\lambda_i(A), 0\} - \max\{\lambda_j(A), 0\}}{\lambda_i(A) - \lambda_j(A)}, \quad i,j = 1,\ldots,n, \tag{6}$$

with the convention that $0/0 := 1$.

2.2 Regularity conditions

In this subsection, we review some fundamental variational properties of optimization problems, which are essential to our subsequent discussions. Most of these can be found in the literature, for example, [36] and [6].

Let us begin with the definition of the tangent cone. For a closed convex set $\mathcal{K} \subseteq \mathbb{S}^n$, the tangent cone to $\mathcal{K}$ at $y \in \mathbb{S}^n$ is defined by

$$\mathcal{T}_{\mathcal{K}}(y) = \big\{d \in \mathbb{S}^n \mid \exists\, t_k \downarrow 0 \text{ such that } \operatorname{dist}(y + t_k d, \mathcal{K}) = o(t_k)\big\}. \tag{7}$$

We remark here that, for a smooth embedded submanifold $\mathcal{M}$ of a Euclidean space $\mathbb{X}$, the tangent cone $\mathcal{T}_{\mathcal{M}}(x)$ to $\mathcal{M}$ at $x \in \mathcal{M}$ coincides with the usual tangent space $\mathcal{T}_x\mathcal{M}$. Nevertheless, we retain both pieces of notation in order to reflect the different traditions in variational analysis and differential geometry. Throughout the paper, the symbol "$\mathcal{T}$" will be used both for the tangent cone $\mathcal{T}_{\mathcal{K}}(y)$ at a point y of a closed convex set $\mathcal{K}$, and for the tangent space $\mathcal{T}_A\mathcal{M}$ at a point A of a manifold $\mathcal{M}$. Although these notions are closely related in spirit, the roles of the set and the point in the notation are interchanged, and this will allow us to keep track of whether we are working in a variational analysis or a differential geometry setting.

Let x be a feasible solution of (1). As noted in [6, Theorem 3.9 and Prop. 3.17], the following Robinson constraint qualification (RCQ) is equivalent to the nonemptiness and boundedness of the corresponding Lagrange multiplier set $M(x)$.

Definition 1 The *Robinson constraint qualification (RCQ)* is said to hold at a feasible point x if

$$g'(x)\mathbb{X} + \mathcal{T}_{\mathbb{S}^n_+}(g(x)) = \mathbb{S}^n.$$

For a stationary point $\overline{x}$ of (1), the strict Robinson constraint qualification condition is defined as follows.

Definition 2 Let $\overline{x} \in \mathbb{X}$ be a stationary point of (1) with a Lagrange multiplier $\overline{y} \in M(\overline{x})$. The *strict Robinson constraint qualification (SRCQ)* is said to hold at $\overline{x}$ with respect to $\overline{y}$ if

$$g'(\overline{x})\mathbb{X} + \mathcal{T}_{\mathbb{S}^n_+}(g(\overline{x})) \cap \overline{y}^{\perp} = \mathbb{S}^n,$$

where $y^{\perp} := \{s \in \mathbb{S}^n \mid \langle s, y\rangle = 0\}$ for any vector $y \in \mathbb{S}^n$.

It follows from [6, Proposition 4.50] that the set of Lagrange multipliers $M(\overline{x})$ is a singleton if the SRCQ holds.

For a feasible point x, the *critical cone* $\mathcal{C}(x)$ of (1) at x is defined by

$$\mathcal{C}(x) = \left\{ v_x \in \mathbb{X} \,\middle|\, g'(x)v_x \in \mathcal{T}_{\mathbb{S}^n_+}(g(x)),\ f'(x)v_x = 0 \right\}. \tag{8}$$

If $\overline{x}$ is a stationary point with $\overline{y} \in M(\overline{x})$, then the critical cone $\mathcal{C}(\overline{x})$ takes the following form:

$$\mathcal{C}(\overline{x}) = \left\{ v_x \in \mathbb{X} \,\middle|\, g'(\overline{x})v_x \in \mathcal{T}_{\mathbb{S}^n_+}(g(\overline{x})) \cap \overline{y}^{\perp} \right\}. \tag{9}$$

The following definition of constraint nondegeneracy for (1), introduced by Robinson [35], is stronger than the SRCQ, since at a KKT pair $(\overline{x}, \overline{y})$ of (1), it follows from [6, Proposition 4.73] that

$$\operatorname{lin}\big(\mathcal{T}_{\mathbb{S}^n_+}(g(\overline{x}))\big) \subseteq \mathcal{T}_{\mathbb{S}^n_+}(g(\overline{x})) \cap \overline{y}^{\perp}.$$

Definition 3 The *constraint nondegeneracy* of (1) is said to hold at a feasible point x if

$$g'(x)\mathbb{X} + \operatorname{lin}(\mathcal{T}_{\mathbb{S}^n_+}(g(x))) = \mathbb{S}^n.$$

Recall that the inner and outer second order tangent sets ([6, (3.49) and (3.50)]) to the given closed set $\mathcal{K} \in \mathbb{Y}$ in the direction $d \in \mathbb{Y}$ are defined, respectively, by

$$\mathcal{T}^{i,2}_{\mathcal{K}}(y, d) := \{w \in \mathbb{Y} \mid \operatorname{dist}(y + td + \frac{1}{2}t^2 w, C) = o(t^2),\ t \geq 0\}$$

and

$$\mathcal{T}^{2}_{\mathcal{K}}(y, d) := \{w \in \mathbb{Y} \mid \exists\, t^k \downarrow 0,\ \operatorname{dist}(y + t^k d + \frac{1}{2}(t^k)^2 w, C) = o((t^k)^2)\}.$$

Note that in general, $\mathcal{T}^{i,2}_{\mathcal{K}}(y, d) \neq \mathcal{T}^{2}_{\mathcal{K}}(y, d)$ even for convex sets $\mathcal{K}$ (cf. [6, Section 3.3]). However, it follows from [6, Proposition 3.136] that if $\mathcal{K}$ is a C^2-cone reducible [6, Definition 3.135] convex set (e.g., the polyhedron, second order cone, positive semidefinite matrix cone and their Cartesian products), then equality always holds. In this case, $\mathcal{T}^{2}_{\mathcal{K}}(y, d)$ will be simply called the second order tangent set to $\mathcal{K}$ at $y \in \mathbb{Y}$ in the direction $d \in \mathbb{Y}$.

The following second order sufficient condition is adopted from [6, (3.276)].

Definition 4 Let $\overline{x} \in \mathbb{X}$ be a stationary point of (1) with a Lagrange multiplier $\overline{y} \in M(\overline{x})$. The *second order sufficient condition (SOSC)* is said to hold at $(\overline{x}, \overline{y})$ if

$$\langle v_x, \nabla^2_{xx} L(\overline{x}, \overline{y}) v_x \rangle - \sigma(\overline{y}, \mathcal{T}^2_{\mathbb{S}^n_+}(g(\overline{x}), g'(\overline{x})v_x)) > 0 \quad \forall\, v_x \in \mathcal{C}(\overline{x}) \setminus \{0\}, \tag{10}$$

where σ denotes the support function.

The following definition of the strong second order sufficient condition for (1) is an extension of [39, Definition 3.2] for the (1).

Definition 5 Let $\overline{x} \in \mathbb{X}$ be a stationary point of (1) with a Lagrange multiplier $\overline{y} \in M(\overline{x})$. The *strong second order sufficient condition (S-SOSC)* is said to hold at $(\overline{x}, \overline{y})$ if

$$\langle v_x, \nabla^2_{xx} L(\overline{x}, \overline{y}) v_x \rangle - \sigma(\overline{y}, \mathcal{T}^2_{\mathbb{S}^n_+}(g(\overline{x}), g'(\overline{x})v_x)) > 0 \quad \forall\, v_x \in \mathrm{app}(\overline{x}, \overline{y}) \setminus \{0\},$$

where $\mathrm{app}(\overline{x}, \overline{y})$ is the outer approximation of the affine hull of the critical cone $\mathcal{C}(\overline{x})$ of (1) with respect to $(\overline{x}, \overline{y})$, i.e.,

$$\mathrm{app}(\overline{x}, \overline{y}) := \{v_x \in \mathbb{X} \mid g'(\overline{x})v_x \in \mathrm{aff}(\mathcal{T}_{\mathbb{S}^n_+}(g(\overline{x})) \cap \overline{y}^\perp)\}. \tag{11}$$

The above definitions of constraint qualifications and second order conditions are classical results in optimization theory. At a local optimum of (1) where the RCQ holds, it is known that the S-SOSC and the constraint nondegeneracy are equivalent to the strong regularity [39], and that the SRCQ and the SOSC are equivalent to the strong metric subregularity [13]. However, it is shown in [19] that there exist weaker constraint qualifications that do not characterize the nonsingularity of the entire generalized Jacobian set, yet still possess advantages for algorithm design. As weaker extensions of the SRCQ condition (Definition 2) and the SOSC condition (Definition 4), respectively, the following weak strict Robinson constraint qualification (W-SRCQ) and weak second order condition (W-SOC) were first introduced in [19] for (1).

Definition 6 Let $\overline{x} \in \mathbb{X}$ be a stationary point of (1) with a Lagrange multiplier $\overline{y} \in M(\overline{x})$. The *weak strict Robinson constraint qualification (W-SRCQ)* is said to hold at $\overline{x}$ with respect to $\overline{y}$ if

$$g'(\overline{x})\mathbb{X} + \mathrm{aff}(\mathcal{T}_{\mathbb{S}^n_+}(g(\overline{x})) \cap \overline{y}^\perp) = \mathbb{S}^n. \tag{12}$$

Definition 7 Let $\overline{x} \in \mathbb{X}$ be a stationary point of (1) with a Lagrange multiplier $\overline{y} \in M(\overline{x})$. The *weak second order condition (W-SOC)* is said to hold at $(\overline{x}, \overline{y})$ if

$$\langle v_x, \nabla^2_{xx} L(\overline{x}, \overline{y}) v_x \rangle - \sigma(\overline{y}, \mathcal{T}^2_{\mathbb{S}^n_+}(g(\overline{x}), g'(\overline{x})v_x)) \neq 0 \quad \forall\, v_x \in \mathrm{appl}(\overline{x}, \overline{y}) \setminus \{0\}, \tag{13}$$

where $\mathrm{appl}(\overline{x}, \overline{y})$ is defined by

$$\mathrm{appl}(\overline{x}, \overline{y}) := \left\{v_x \in \mathbb{X} \mid g'(\overline{x})v_x \in \mathrm{lin}(\mathcal{T}_{\mathbb{S}^n_+}(g(\overline{x})) \cap \overline{y}^\perp)\right\}. \tag{14}$$

Remark 2 Here, the definition of W-SOC is modified from "> 0" in [19] to "$\neq 0$". This change is made to achieve a more elegant equivalence, which will be established in Theorem 2. Moreover, by a continuity argument, it is straightforward to show that "$\neq 0$" implies the quadratic form maintains a consistent sign (either positive or negative). Furthermore, under assumptions such as (1) being a convex programming or the second order necessary condition holding, our definition is equivalent to the original "> 0" condition and thus consistent with that in [19].

We refer to the constraint qualifications and second order conditions in Definitions 1–7 as *problem-level* regularity conditions, as they can be verified directly from the problem data. In contrast, the following conditions are called *solution-level* regularity conditions, since they require solving equations and computing distances, tasks that are generally difficult in practice. It should be noted that the solution-level conditions correspond to the conventional notion of regularity in variational analysis. Nevertheless, we also regard constraint qualifications and second order conditions as regularity conditions, because, as our stratified perspective shows, they encode the geometric and topological structure that ensures well-behaved local behavior around the solution.

The following strong metric regularity is the most well-known one among solution-level regularity conditions.

Definition 8 We say $\Phi : \mathbb{X} \rightrightarrows \mathbb{Y}$ is *strongly metrically regular* at $(\overline{x}, \overline{y})$, if $\overline{y} \in \Phi(\overline{x})$, and there exist constants $\kappa > 0$ and neighborhoods $\mathcal{U}$ of $\overline{x}$, $\mathcal{V}$ of $\overline{y}$, such that for all $x \in \mathcal{U}$ and $y \in \mathcal{V}$, the following inequality holds

$$\mathrm{dist}(x, \Phi^{-1}(y)) \leq \kappa \cdot \mathrm{dist}(y, \Phi(x))$$

and $\Phi^{-1}(y) \cap \mathcal{U}$ is a singleton for every $y \in \mathcal{V}$.

It is well known that, at a KKT pair of (1), strong metric regularity of the KKT natural mapping F we considered is equivalent to the strong regularity [35] of the KKT system formulated as a generalized equation [15]. As stated in the introduction, at a local optimum where the RCQ holds, the above two conditions are further equivalent to the simultaneous satisfaction of the S-SOSC (Definition 5) and the constraint nondegeneracy (Definition 3).

Metric regularity is a weaker property, as it does not require the inverse mapping to be single-valued.

Definition 9 We say $\Phi : \mathbb{X} \rightrightarrows \mathbb{Y}$ is *metrically regular* at $(\overline{x}, \overline{y})$ if $\overline{y} \in \Phi(\overline{x})$, and there exist a constant $\kappa > 0$ and neighborhoods $\mathcal{U}$ of $\overline{x}$, $\mathcal{V}$ of $\overline{y}$, such that for all $x \in \mathcal{U}$ and $y \in \mathcal{V}$, the following inequality holds:

$$\operatorname{dist}(x, \Phi^{-1}(y)) \leq \kappa \cdot \operatorname{dist}(y, \Phi(x)).$$

Although metric regularity is, in general, weaker than strong metric regularity, the two notions coincide in a number of important settings. In particular, Ma et al. [29] show that this equivalence holds for the NLSDP (1) considered here.

Another closely related notion is the *local error bound*. Various error bound formulations appear in the literature, and a systematic discussion is beyond the scope of this paper. We record one standard version below, noting that it follows directly from metric regularity.

Definition 10 Let $F : \mathbb{X} \to \mathbb{Y}$ be a single-valued mapping, and let $\overline{x} \in \mathbb{X}$ satisfy $F(\overline{x}) = 0$. We say that the equation $F(x) = 0$ admits a *local error bound* at $\overline{x}$ if there exist a constant $\kappa > 0$ and a neighborhood $\mathcal{U}$ of $\overline{x}$ such that for all $x \in \mathcal{U}$,

$$\operatorname{dist}\big(x, F^{-1}(0)\big) \leq \kappa \cdot \|F(x)\|.$$

2.3 Manifolds and stratification

In this subsection, we briefly recall some notions from differential geometry and stratification theory that will be used throughout the paper. We do not reproduce the basic definitions of manifolds, local charts, tangent spaces, tangent maps, tangent bundles, Riemannian metrics, or exponential maps, and instead refer the reader to standard texts such as [27,1] for details.

First, we should address a potential ambiguity in notation. Suppose that $\mathcal{M}$ is a manifold embedded in a Euclidean space $\mathbb{Y}$, and let $p \in \mathcal{M}$. Via the canonical identification $\mathcal{T}_p\mathbb{Y} \cong \mathbb{Y}$, we may regard any tangent vector $v \in \mathcal{T}_p\mathbb{Y}$ as an element of $\mathbb{Y}$ itself, and thus form the sum $p + v$ using the ambient vector space structure of $\mathbb{Y}$. Moreover, we use the notation dh_x to denote the differential of a mapping between manifolds. For mappings on Euclidean space, although dh_x is also standard for the differential, we will instead use $h'(x)$ to help readers distinguish whether we are working on manifolds or in Euclidean space. Furthermore, the notation $\nabla h(x) := \big(h'(x)\big)^*$ is used to denote the adjoint of the linear operator $h'(x)$, in accordance with optimization conventions.

We next recall the notion of transversality from differential topology (see, e.g., [23, Chapter 3.2], which extends the transversality concept used in variational analysis [6]. This notion formalizes the idea that a smooth map meets a given submanifold in a "regular" way.

Definition 11 [23, Chapter 3.2] Suppose $h : \mathcal{M} \to \mathcal{N}$ is a smooth map between smooth manifolds $\mathcal{M}$ and $\mathcal{N}$, $\mathcal{N}' \subseteq \mathcal{N}$ is a submanifold of $\mathcal{N}$. Then we say h *intersects with* $\mathcal{N}'$ *transversally* at x in $\mathcal{N}$, or simply h *is transverse to* $\mathcal{N}'$ at x in $\mathcal{N}$, denoted by $h \pitchfork_x \mathcal{N}'$ in $\mathcal{N}$, whenever $h(x) \in \mathcal{N}'$,

$$dh_x(\mathcal{T}_x\mathcal{M}) + \mathcal{T}_{h(x)}\mathcal{N}' = \mathcal{T}_{h(x)}\mathcal{N}, \tag{15}$$

where dh_x is the differential of h at x and $\mathcal{T}$ denotes the tangent space. Moreover, if (15) holds whenever $x \in \mathcal{L}$ and $h(x) \in \mathcal{N}'$, where $\mathcal{L}$ is a closed subset in $\mathcal{M}$, then we say h *intersects with* $\mathcal{N}'$ *transversally along* $\mathcal{L}$ in $\mathcal{N}$, denoted by $h \pitchfork_{\mathcal{L}} \mathcal{N}'$ in $\mathcal{N}$. And we also just denote $h \pitchfork \mathcal{N}'$ in $\mathcal{N}$ if $\mathcal{L} = \mathcal{M}$.

Stratification is a fundamental notion in topology and singularity theory. It is well-known that smooth manifolds furnish a flexible framework that covers many nonlinear spaces, such as orthogonal groups and Grassmannians. However, sets with singularities lie outside the class of manifolds. The concept of a stratified space is designed as a natural extension of the manifold framework to accommodate such singular sets, for instance, real algebraic varieties [43,44], which arise frequently in polynomial optimization and related areas. For background on (Whitney) stratifications, we refer the reader to [43,44,30,31]. In the present work, we only require a basic version of the theory, and the following definition will be sufficient for our subsequent developments.

Definition 12 Let $\mathbb{X}$ be a Euclidean space and $\mathcal{X} \subseteq \mathbb{X}$ a closed subset. A *stratification* of $\mathcal{X}$ is a locally finite partition $S = \{\mathcal{S}_i\}_{i \in I}$ of $\mathcal{X}$ into connected embedded C^k submanifolds (typically $k \geq 1$) $\mathcal{S}_i$ (called *strata*) such that

1. *Frontier condition*: For any two strata $\mathcal{S}_i, \mathcal{S}_j \in S$, if $\mathcal{S}_i \cap \overline{\mathcal{S}_j} \neq \emptyset$, then $\mathcal{S}_i \subseteq \overline{\mathcal{S}_j}$, where $\overline{\mathcal{S}_j}$ means the closure of $\mathcal{S}_j$. This implies the boundary $\overline{\mathcal{S}_j} \setminus \mathcal{S}_j$ is a union of strata.
2. *Local finiteness*: Every point $x \in \mathcal{X}$ has a neighborhood in $\mathbb{X}$ intersecting finitely many strata.

A topological space with a stratification structure is called a *stratified space*.

Roughly speaking, a stratification may be viewed as a decomposition of a given set into smooth pieces together with their well-organized boundaries. In Section 3, we shall introduce a stratification of the Euclidean space $\mathbb{S}^n$ tailored to our purposes, even though $\mathbb{S}^n$ is itself a linear space and hence a smooth manifold without singularities.

3 The index stratification of $\mathbb{S}^n$

In this section, we study a stratification of $\mathbb{S}^n$, which decomposes it into finitely many smooth strata corresponding to matrices with prescribed numbers of positive and negative eigenvalues (see (17)). We refer to this decomposition as the *index stratification* of $\mathbb{S}^n$. It is naturally adapted to the positive semidefinite cone $\mathbb{S}^n_+$, since $\mathbb{S}^n_+$ is the union of a subcollection of these strata and the projection mapping onto $\mathbb{S}^n_+$ is smooth when restricted to each individual stratum.

The symmetric matrix space $\mathbb{S}^n$ is linearly isometric to $\mathbb{R}^{\frac{n(n+1)}{2}}$ and therefore admits many stratifications in the sense of Definition 12. In what follows, we single out a particular stratification that is useful for our analysis. Although this decomposition has appeared implicitly in earlier work, we record it here together with a concise description of its basic geometric properties.

For given integers $p, q \in \mathbb{Z}_+$ with $p + q \leq n$, define

$$\mathcal{M}_{p,q} := \big\{A \in \mathbb{S}^n \mid |\alpha(A)| = p, \quad |\gamma(A)| = q\big\}, \tag{16}$$

and, when no confusion can arise, write $\mathcal{M}$ in place of $\mathcal{M}_{p,q}$. It is shown in [22] that each $\mathcal{M}_{p,q}$ is a smooth embedded submanifold of $\mathbb{S}^n$ of dimension $\dim(\mathcal{M}_{p,q}) = n(p+q) - \frac{1}{2}(p+q)(p+q-1)$. It is clear that the family $\{\mathcal{M}_{p,q}\}_{0\leq p+q\leq n}$ yields a decomposition of $\mathbb{S}^n$ into a finite disjoint union, that is,

$$\mathbb{S}^n = \bigcup_{0\leq p+q\leq n} \mathcal{M}_{p,q}. \tag{17}$$

It is easy to check that this decomposition satisfies the frontier condition and the local finiteness given in Definition 12. Hence it forms a stratification of $\mathbb{S}^n$, which we shall refer to as the *index stratification* of $\mathbb{S}^n$[1], and each $\mathcal{M}_{p,q}$ will be called a *stratum* of $\mathbb{S}^n$.

Figure 1 illustrates the stratification of $\mathbb{S}^2 \simeq \mathbb{R}^3$. This space decomposes into three 3-dimensional strata, two 2-dimensional strata, and the trivial zero-dimensional stratum $\{0\}$. Moreover, the adjacency among these strata satisfies the property that the closure of any stratum is precisely the union of itself and all strata of strictly lower dimension, in accordance with the frontier condition in Definition 12.

Before examining further properties of the index stratification of $\mathbb{S}^n_+$, we introduce a convenient representation of the tangent space $\mathcal{T}_A\mathcal{M}_{p,q}$ of a stratum $\mathcal{M}_{p,q}$ at a point $A \in \mathcal{M}_{p,q}$. It is well-known (see, e.g., [22]) that

$$\mathcal{T}_A\mathcal{M}_{p,q} = \big\{H = X^\top A + AX \in \mathbb{S}^n \mid X \in \mathbb{R}^{n\times n}\big\}. \tag{18}$$

The next result provides an alternative characterization in terms of the eigenvalue decomposition (4).

Proposition 1 *Let $A \in \mathbb{S}^n$ have an indexed eigenvalue decomposition (IED) $(\alpha, \beta, \gamma, p, q, P, \lambda)$. Then, we have*

$$\mathcal{T}_A\mathcal{M}_{p,q} = \big\{H \in \mathbb{S}^n \mid (P^\top HP)_{\beta\beta} = 0\big\}. \tag{19}$$

Proof Denote the right-hand side of (19) by $\mathcal{R}$. Let $H \in \mathcal{T}_A\mathcal{M}_{p,q}$ be arbitrarily given. It then follows from (18) that there exists $X \in \mathbb{R}^{n\times n}$ such that $H = X^\top A + AX$. Denote $\widetilde{H} := P^\top HP$. We then have

$$\widetilde{H} = P^\top X^\top P \begin{bmatrix} \Lambda_{\alpha\alpha} & 0 & 0 \\ 0 & \Lambda_{\beta\beta} & 0 \\ 0 & 0 & \Lambda_{\gamma\gamma} \end{bmatrix} + \begin{bmatrix} \Lambda_{\alpha\alpha} & 0 & 0 \\ 0 & \Lambda_{\beta\beta} & 0 \\ 0 & 0 & \Lambda_{\gamma\gamma} \end{bmatrix} P^\top XP,$$

[1] In fact, the family $\{\mathcal{M}_{p,q}\}_{0\leq p+q\leq n}$ forms a Whitney stratification of $\mathbb{S}^n$ (see, for example, the definition of Whitney stratifications in [43,44,30]). This follows from the canonical Whitney stratification of the real algebraic variety of symmetric matrices with prescribed inertia, defined by the vanishing of suitable minors; we refer to [43,44] for a rigorous treatment. Moreover, Olikier [32] shows that the index stratification of $\mathbb{S}^n_+$ satisfies certain regularity properties that are particularly suited for first-order optimality analysis.

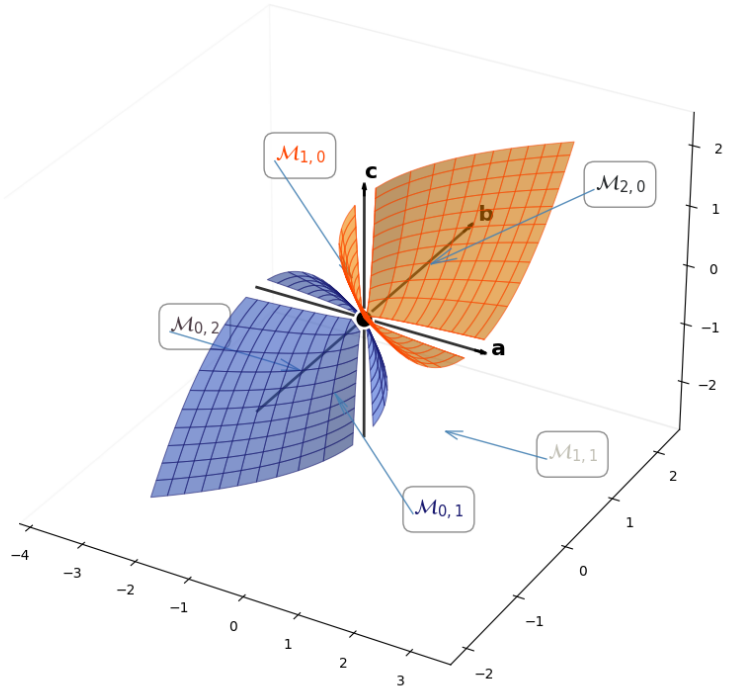

Fig. 1: The index stratification for $\mathbb{S}^2$

which, together with $\Lambda_{\beta\beta} = 0$, yields that $\widetilde{H}_{\beta\beta} = 0$. Thus, we have $\mathcal{T}_A\mathcal{M}_{p,q} \subseteq \mathcal{R}$.

On the other hand, it is clear that the dimension of the subspace $\mathcal{R}$ is $\dim \mathcal{R} = n(p+q) - \frac{1}{2}(p+q)(p+q-1)$. Moreover, it follows from [22, Proposition 2.1] that $\dim \mathcal{T}_A\mathcal{M}_{p,q} = n(p+q) - \frac{1}{2}(p+q)(p+q-1)$. Thus, we conclude that $\mathcal{T}_A\mathcal{M}_{p,q} = \mathcal{R}$. □

With the stratified structure and the characterization of the tangent spaces in hand, we can now analyze the differentiability of the projection operator $\Pi_{\mathbb{S}^n_+}$ along the strata. In particular, we shall see that, when restricted to a fixed stratum $\mathcal{M}_{p,q}$, the mapping $\Pi_{\mathbb{S}^n_+} : \mathcal{M}_{p,q} \to \mathcal{M}_{p,0}$ is not only differentiable with an explicit closed-form differential, but is in fact of class C^∞, thus strengthening the previously known C^1-differentiability (see Remark 4). These regularity properties are fundamental for the variational analysis and the algorithmic developments in our subsequent discussions. Since the proof relies on analytic functional calculus for symmetric matrices, it is deferred to Appendix A.

Theorem 1 *For any $\mathcal{M} = \mathcal{M}_{p,q}$, the restricting of $\Pi_{\mathbb{S}^n_+}$ on $\mathcal{M}$, denoted by $\Pi_{\mathbb{S}^n_+}|_{\mathcal{M}} : \mathcal{M} \to \mathbb{S}^n$, is a C^∞ map between smooth manifolds $\mathcal{M}$ and $\mathbb{S}^n$. Moreover, given an arbitrary $A \in \mathcal{M}$ with an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$, the differential $\xi_A := d(\Pi_{\mathbb{S}^+_n}|_{\mathcal{M}})_A : \mathcal{T}_A\mathcal{M} \to \mathcal{T}_{\Pi_{\mathbb{S}^n_+}(A)}\mathbb{S}^n$ is given by, for $H \in \mathcal{T}_A\mathcal{M}$,*

$$\xi_A(H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \Xi_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}^\top_{\alpha\beta} & 0 & 0 \\ \widetilde{H}^\top_{\alpha\gamma} \circ \Xi^\top_{\alpha\gamma} & 0 & 0 \end{bmatrix} P^\top, \tag{20}$$

where $\widetilde{H} = P^\top H P$.

Remark 3 It is straightforward to verify that $\Pi_{\mathbb{S}^n_+}$ maps $\mathcal{M}_{p,q}$ onto $\mathcal{M}_{p,0}$, which is smoothly embedded in $\mathbb{S}^n$. Since we have already shown that $\Pi_{\mathbb{S}^n_+} : \mathcal{M}_{p,q} \to \mathbb{S}^n$ is a C^∞ map, standard properties of smooth embeddings imply that $\Pi_{\mathbb{S}^n_+} : \mathcal{M}_{p,q} \to \mathcal{M}_{p,0}$ is also C^∞. In the sequel, we shall omit the explicit restriction notation $|_{\mathcal{M}_{p,q}}$ and simply write $\Pi_{\mathbb{S}^n_+} : \mathcal{M}_{p,q} \to \mathbb{S}^n$ or $\Pi_{\mathbb{S}^n_+} : \mathcal{M}_{p,q} \to \mathcal{M}_{p,0}$, depending on the context.

Remark 4 The C^∞-smoothness established in Theorem 1, together with the explicit closed-form expression of the differential in (20), refines and strengthens the existing results for the projector $\Pi_{\mathbb{S}^n_+}$ along the strata. We emphasize that the C^1-differentiability of $\Pi_{\mathbb{S}^n_+}$ relative to a stratum $\mathcal{M}_{p,q}$ can already be derived by known

arguments, for instance via the Löwner functional calculus or the general theory of C^1-partly smooth mappings (see, e.g., [28, Example 4.14] and [17, Proposition 9.7]). Theorem 1 goes one step further by establishing full C^∞-smoothness and by providing an explicit formula for the differential, a result which, to the best of our knowledge, has not been available in the literature.

Recall that the nonsmoothness of $F(z)$ in (2) stems solely from the projection term $\Pi_{\mathbb{S}^n_+}\big(G(z)\big)$. Our goal is therefore to isolate and exploit the smooth structure that emerges once the positive and negative inertia indices of $G(z)$ are fixed. To this end, we lift the index stratification of $\mathbb{S}^n$ to the product space $\mathbb{X}\times\mathbb{S}^n$ via the mapping G. For each pair (p,q) with $0\le p+q\le n$, define

$$\widetilde{\mathcal{M}}_{p,q} := G^{-1}(\mathcal{M}_{p,q}) := \{z=(x,y)\in\mathbb{X}\times\mathbb{S}^n \mid G(z)=g(x)+y\in\mathcal{M}_{p,q}\}. \tag{21}$$

Since G is a smooth submersion (see, e.g., [27, Chapter 4]), it follows from [27, Corollary 6.31] that each $\widetilde{\mathcal{M}}_{p,q}$ is a smoothly embedded submanifold of $\mathbb{X}\times\mathbb{S}^n$. Moreover, one readily verifies that

$$\mathbb{X}\times\mathbb{S}^n = \bigcup_{0\le p+q\le n}\widetilde{\mathcal{M}}_{p,q}$$

is a stratification of $\mathbb{X}\times\mathbb{S}^n$, which we continue to call the *index stratification*. When the indices p and q are clear from the context, we shall simply write $\widetilde{\mathcal{M}}$ in place of $\widetilde{\mathcal{M}}_{p,q}$.

By Theorem 1, although F is nonsmooth on $\mathbb{X}\times\mathbb{S}^n$, it becomes smooth when restricted to any fixed stratum $\widetilde{\mathcal{M}}_{p,q}$. The following lemma, obtained by a direct computation, characterizes the tangent space of $\widetilde{\mathcal{M}}_{p,q}$ and will be used to simplify the calculation of the differential of F along $\widetilde{\mathcal{M}}_{p,q}$. For brevity, we omit the proof.

Lemma 1 *Fix (p,q) with $0\le p+q\le n$, and let $\mathcal{M}:=\mathcal{M}_{p,q}$ and $\widetilde{\mathcal{M}}:=\widetilde{\mathcal{M}}_{p,q}=G^{-1}(\mathcal{M})$. Take any $z=(x,y)\in\widetilde{\mathcal{M}}$. Then the tangent space $\mathcal{T}_z\widetilde{\mathcal{M}}$ can be represented as*

$$\mathcal{T}_z\widetilde{\mathcal{M}} = \big\{(v_x,v_y)\in\mathbb{X}\times\mathbb{S}^n \mid \nabla g(x)^*v_x+v_y\in\mathcal{T}_{G(z)}\mathcal{M}\big\}. \tag{22}$$

In particular, the linear mapping $\phi_z:\mathcal{T}_z\widetilde{\mathcal{M}}\to\mathbb{X}\times\mathcal{T}_{G(z)}\mathcal{M}$ given by

$$\phi_z(v_x,v_y) := \big(v_x,\ \nabla g(x)^*v_x+v_y\big) \tag{23}$$

is a linear isomorphism with inverse $\phi_z^{-1}:\mathbb{X}\times\mathcal{T}_{G(z)}\mathcal{M}\to\mathcal{T}_z\widetilde{\mathcal{M}}$ given by

$$\phi_z^{-1}(v_x,H) := \big(v_x,\ H-\nabla g(x)^*v_x\big). \tag{24}$$

In the following, we write $v\in\mathcal{T}_z\widetilde{\mathcal{M}}$ as (v_x,v_y) if we regard $v\in\mathcal{T}_z\widetilde{\mathcal{M}}\subseteq\mathbb{X}\times\mathbb{S}^n$ and write (v_x,H) if we express v under the isomorphism ϕ_z. Moreover, the assignments $z\mapsto\phi_z$ and $z\mapsto\phi_z^{-1}$ are smooth (in the usual sense of smooth maps into the space of linear operators).

By the chain rule, together with the expression for the differential of $\Pi_{\mathbb{S}^n_+}$ restricted to $\mathcal{M}$ given in Theorem 1, Lemma 1 yields an explicit representation of the differential of F along $\widetilde{\mathcal{M}}$. More precisely, for $z=(x,y)\in\widetilde{\mathcal{M}}$ and $(v_x,v_y)\in\mathcal{T}_z\widetilde{\mathcal{M}}$, set

$$H := \nabla g(x)^*v_x+v_y\in\mathcal{T}_{G(z)}\mathcal{M}.$$

Then, for any $v=(v_x,v_y)\in\mathcal{T}_z\widetilde{\mathcal{M}}$,

$$\begin{aligned} dF_z v &= \begin{bmatrix}\nabla^2_{xx}L(z)\,v_x+\nabla g(x)\,v_y\\ -\nabla g(x)^*v_x+\xi_{G(z)}\big(dG(z)(v_x,v_y)\big)\end{bmatrix} = \begin{bmatrix}\nabla^2_{xx}L(z)\,v_x+\nabla g(x)\big(H-\nabla g(x)^*v_x\big)\\ -\nabla g(x)^*v_x+\xi_{G(z)}(H)\end{bmatrix}\\ &= \begin{bmatrix}\big(\nabla^2_{xx}L(z)-\nabla g(x)\nabla g(x)^*\big)v_x+\nabla g(x)\,H\\ -\nabla g(x)^*v_x+\xi_{G(z)}(H)\end{bmatrix}. \end{aligned} \tag{25}$$

Equivalently, with respect to the coordinates $(v_x,H)\in\mathbb{X}\times\mathcal{T}_{G(z)}\mathcal{M}$ induced by (23), the above identity can be written in the block form

$$dF_z(v_x,v_y) = \begin{bmatrix}\nabla^2_{xx}L(z)-\nabla g(x)\nabla g(x)^* & \nabla g(x)\\ -\nabla g(x)^* & \xi_{G(z)}\end{bmatrix}\begin{bmatrix}v_x\\ H\end{bmatrix}. \tag{26}$$

The remainder of this section is divided into two subsections. We first carry out a variational analysis on a fixed stratum, and then investigate how the relevant regularity properties behave when moving across different strata. Together, these developments provide the theoretical foundation for designing algorithms that achieve global convergence and fast local convergence rates, relying on weaker conditions than the classical regularity conditions.

3.1 Variational properties on stratum

In this subsection, we first characterize the *stratum-restricted strong metric regularity* (Definition 13) on a fixed stratum. To describe the behavior of the problem along directions confined to that stratum, we then analyze a related transversality condition, which yields a geometric interpretation of the proposed regularity framework.

3.1.1 The stratum-restricted strong metric regularity

The following variational property is a natural stratum-restricted extension of strong metric regularity (Definition 8). It is worth noting that the proposed property is much weaker than its classical counterpart defined on the entire space, especially in degenerate cases.

Definition 13 Let $\mathbb{X}$ be a stratified space. We say $\Phi : \mathbb{X} \rightrightarrows \mathbb{Y}$ is *stratum-restricted strongly metric regular* on a stratum $\mathcal{M} \subseteq \mathbb{X}$ at $(\overline{x}, \overline{y})$, if $\overline{x} \in \mathcal{M}$, $\overline{y} \in \Phi(\overline{x})$, and there exist constants $\kappa > 0$ and neighborhoods $\mathcal{U}$ of $\overline{x}$, $\mathcal{V}$ of $\overline{y}$, such that for all $x \in \mathcal{U} \cap \mathcal{M}$ and $y \in \mathcal{V} \cap \Phi(\mathcal{M})$, the following inequality holds:

$$\operatorname{dist}(x, \Phi^{-1}(y) \cap \mathcal{M}) \leq \kappa \cdot \operatorname{dist}(y, \Phi(x)),$$

and $(\Phi^{-1}(y) \cap \mathcal{M}) \cap \mathcal{U}$ is a singleton for every $y \in \mathcal{V} \cap \Phi(\mathcal{M})$.

Remark 5 In our setting, since F is single-valued and smooth on $\widetilde{\mathcal{M}}$, for any $\overline{z} \in \widetilde{\mathcal{M}}$, the stratum-restricted strong metric regularity of F on $\widetilde{\mathcal{M}}$ at $(\overline{z}, F(\overline{z}))$ is equivalent to the restriction $F|_{\widetilde{\mathcal{M}}}$ being a local Lipschitz homeomorphism near $\overline{z}$.

In the classical theory for (1), strong metric regularity is typically analyzed under the strong second order sufficient condition together with constraint nondegeneracy. Here, we take a parallel route, using the weak second order condition (W-SOC) and the weak strict Robinson constraint qualification (W-SRCQ) instead as our starting point. Accordingly, we first present several equivalent characterizations of these conditions that are more amenable to verification and will be convenient for the subsequent analysis. The proof is straightforward, and we omit it here; interested readers are referred to [39, Lemma 3.1] and [19, Lemma 1].

Lemma 2 *Let $\overline{z} = (\overline{x}, \overline{y})$ be a KKT pair of (1) with an IED $(\alpha, \beta, \gamma, p, q, \overline{P}, \overline{\lambda})$ of $G(\overline{z})$. Then there hold*

- *(a) The W-SOC (Definition 7) holds at $\overline{z}$ if and only if for all $v_x \in \operatorname{appl}(\overline{z}) \setminus \{0\}$,*

$$\left\langle v_x, \nabla^2_{xx} L(\overline{z}) v_x \right\rangle - \sigma_{\overline{z}}(v_x) \neq 0,$$

where $\operatorname{appl}(\overline{z})$ is defined at (14) and

$$\sigma_{\overline{z}}(v_x) = 2 \sum_{i \in \alpha} \sum_{j \in \gamma} \frac{\overline{\lambda}_j}{\overline{\lambda}_i} \left[\overline{P}^\top (\nabla g(\overline{x})^* v_x) \overline{P} \right]^2_{ij}.$$

Moreover, the set $\operatorname{appl}(\overline{z})$ can be characterized as

$$\operatorname{appl}(\overline{z}) = \left\{ v_x \in \mathbb{X} \mid [\overline{P}^\top (\nabla g(\overline{x})^* v_x) \overline{P}]_{\beta\beta} = 0,\ [\overline{P}^\top (\nabla g(\overline{x})^* v_x) \overline{P}]_{\beta\gamma} = 0,\ [\overline{P}^\top (\nabla g(\overline{x})^* v_x) \overline{P}]_{\gamma\gamma} = 0 \right\}. \quad (27)$$

- *(b) The W-SRCQ (Definition 6) holds at $\overline{z}$ if and only if the following implication is satisfied*

$$g'(\overline{x}) \mathbb{X} + \left\{ \overline{P} B \overline{P}^\top \mid B \in \mathbb{S}^n,\ B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0 \right\} = \mathbb{S}^n. \quad (28)$$

Motivated by Lemma 2, we introduce extensions of the W-SOC and the W-SRCQ that apply also at non-KKT pairs. This allows us to investigate perturbation properties of these problem-level regularity conditions in a unified manner.

Definition 14 Considering (1), for $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$ with an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$ of $G(z)$, the weak second order condition (W-SOC) is said to hold at z if

$$\langle v_x, \nabla^2_{xx} L(x, y) v_x \rangle + 2 \sum_{i \in \alpha} \sum_{j \in \gamma} \frac{-\lambda_j}{\lambda_i} \left[P^\top (\nabla g(x)^* v_x) P \right]^2_{ij} \neq 0 \quad \forall\, v_x \in \operatorname{appl}(x, y) \setminus \{0\}, \quad (29)$$

where $\operatorname{appl}(x, y)$ is given by

$$\operatorname{appl}(z) = \left\{ v_x \in \mathbb{X} \mid [P^\top (\nabla g(x)^* v_x) P]_{\beta\beta} = 0,\ [P^\top (\nabla g(x)^* v_x) P]_{\beta\gamma} = 0,\ [P^\top (\nabla g(x)^* v_x) P]_{\gamma\gamma} = 0 \right\}. \quad (30)$$

Definition 15 Considering (1), for $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$ with an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$ of $G(z)$, the weak strict Robinson constraint qualification (W-SRCQ) is said to hold at z if

$$g'(x)\mathbb{X} + \left\{PBP^\top \mid B \in \mathbb{S}^n,\ B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0\right\} = \mathbb{S}^n. \tag{31}$$

By Lemma 2, one readily verifies that, when z is a KKT pair, Definitions 14 and 15 reduce to the W-SOC in Definition 7 and the W-SRCQ in Definition 6, respectively. This extension enables a unified treatment of both KKT and non-KKT pairs within the same framework.

Meanwhile, we also introduce extensions of the classical problem-level regularity conditions, the S-SOSC and the constraint nondegeneracy, that apply at non-KKT pairs. These are not used in this section, but serve as preparation for the discussion in Section 3.2.

Definition 16 Considering (1), for $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$ with an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$ of $G(z)$, the strong second order sufficient condition (S-SOSC) is said to hold at z if

$$\langle v_x, \nabla^2_{xx} L(x, y) v_x\rangle + 2\sum_{i\in\alpha}\sum_{j\in\gamma} \frac{-\lambda_j}{\lambda_i}\left[P^\top(\nabla g(x)^* v_x)P\right]^2_{ij} > 0 \quad \forall\, v_x \in \mathrm{app}(x, y) \setminus \{0\},$$

where $\mathrm{app}(x, y)$ is given by

$$\mathrm{app}(z) := \left\{v_x \in \mathbb{X} \mid [P^\top(\nabla g(x)^* v_x)P]_{\beta\gamma} = 0,\ [P^\top(\nabla g(x)^* v_x)P]_{\gamma\gamma} = 0\right\}.$$

Definition 17 Considering (1), for $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$ with an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$ of $G(z)$, the constraint nondegeneracy is said to hold at z if

$$g'(x)\mathbb{X} + \left\{PBP^\top \mid B \in \mathbb{S}^n,\ B_{\beta\beta} = 0,\ B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0\right\} = \mathbb{S}^n. \tag{32}$$

Similarly to Lemma 2, one can verify that when z is a KKT pair, Definition 16 and Definition 17 reduce to the S-SOSC in Definition 5 and the constraint nondegeneracy in Definition 3, respectively. In particular, the constraint nondegeneracy stated in Definition 3 at a feasible point x is equivalent to the constraint nondegeneracy in Definition 17 evaluated at the primal-dual pair (x, y), where y satisfies the complementarity condition $\mathbb{S}^n_+ \ni g(x) \perp y \in \mathbb{S}^n_-$.

From now on, whenever we refer to S-SOSC or constraint nondegeneracy, we always mean those given in Definition 16 and Definition 17, unless otherwise specified.

With the characterizations of the W-SOC and the W-SRCQ, and with the differential of the KKT mapping F along $\widetilde{\mathcal{M}}$, we are ready to present the main result of this subsection. The following theorem provides the complete characterization of the *stratum-restricted strong metric regularity* of F on $\widetilde{\mathcal{M}}$.

Theorem 2 *Let $\overline{z} = (\overline{x}, \overline{y})$ with an IED $(\alpha, \beta, \gamma, p, q, \overline{P}, \overline{\lambda})$ of $G(\overline{z})$. Consider the following three statements:*

(a) The W-SOC and the W-SRCQ holds at $\overline{z}$;
(b) The manifold differential $dF_{\overline{z}} : \mathcal{T}_{\overline{z}}\widetilde{\mathcal{M}} \to \mathbb{X} \times \mathbb{S}^n$ is injective;
(c) F is stratum-restricted strongly metric regular on $\widetilde{\mathcal{M}}$ at $(\overline{z}, F(\overline{z}))$.

Then, (a) is equivalent to (b). Moreover, if $\overline{z}$ is a KKT pair of (1), then they are further equivalent to (c).

Proof First, we show that (a)⇒(b). Assume that $(v_x, v_y) \in \mathcal{T}_z\widetilde{\mathcal{M}}$ satisfies $dF(\overline{z})(v_x, v_y) = 0$. Then, it follows from (26) that

$$\begin{bmatrix} \nabla^2_{xx} L(\overline{z}) - \nabla g(\overline{x})\nabla g(\overline{x})^* & \nabla g(\overline{x}) \\ -\nabla g(\overline{x})^* & \xi_{G(\overline{z})} \end{bmatrix} \begin{bmatrix} v_x \\ H \end{bmatrix} = 0, \tag{33}$$

where $H = \nabla g(\overline{x})^* v_x + v_y \in \mathcal{T}_{G(\overline{z})}\mathcal{M}$, which implies

$$-\nabla g(\overline{x})^* v_x + \xi_{G(\overline{z})} H = 0.$$

Then, we have

$$\begin{aligned}
&(\overline{P}^\top(\nabla g(\overline{x})^* v_x)\overline{P})_{ij} = (\overline{P}^\top H \overline{P})_{ij}, && \text{for } (i, j) \in \alpha\alpha, \alpha\beta, \\
&(\overline{P}^\top(\nabla g(\overline{x})^* v_x)\overline{P})_{ij} = 0, && \text{for } (i, j) \in \beta\beta, \beta\gamma, \gamma\gamma, \\
&(\overline{P}^\top(\nabla g(\overline{x})^* v_x)\overline{P})_{ij} = \overline{\Xi}_{ij}(\overline{P}^\top H \overline{P})_{ij}, && \text{for } (i, j) \in \alpha\gamma,
\end{aligned}$$

where we have used the precise expression of $\xi_{G(\overline{z})}$ from (20) and the matrix $\overline{\Xi}$ is defined in (6). Thus, we conclude that $v_x \in \mathrm{appl}(\overline{z})$.

Next, taking the inner product of the first row with v_x, we find

$$\begin{aligned}
0 =& \langle v_x, \nabla^2_{xx} L(\overline{z}) v_x - \nabla g(\overline{x}) \nabla g(\overline{x})^* v_x + \nabla g(\overline{x}) H \rangle \\
=& \langle v_x, \nabla^2_{xx} L(\overline{z}) v_x \rangle + \langle \nabla g(\overline{x})^* v_x, -\nabla g(\overline{x})^* v_x + H \rangle \\
=& \langle v_x, \nabla^2_{xx} L(\overline{z}) v_x \rangle + \langle \overline{P}^\top (\nabla g(\overline{x})^* v_x) \overline{P}, -\overline{P}^\top (\nabla g(\overline{x})^* v_x) \overline{P} + \overline{P}^\top H \overline{P} \rangle \\
=& \langle v_x, \nabla^2_{xx} L(\overline{z}) v_x \rangle - 2 \sum_{i \in \alpha} \sum_{j \in \gamma} \left(1 - \frac{1}{\overline{\Xi}_{ij}} \right) [\overline{P}^\top (\nabla g(\overline{x})^* v_x) \overline{P}]^2_{ij} \\
=& \langle v_x, \nabla^2_{xx} L(\overline{z}) v_x \rangle - \sigma_{\overline{z}}(v_x).
\end{aligned} \tag{34}$$

By the characterization of the W-SOC in Lemma 2 (a), it follows that $v_x = 0$. Hence, (33) reduces to

$$\begin{cases} \nabla g(\overline{x}) H = 0, \\ -\xi_{G(\overline{z})}(H) = 0, \end{cases} \tag{35}$$

which then implies $H = 0$ by the characterization of the W-SRCQ in Lemma 2 (b). Therefore, we know that $dF_{\overline{z}}$ is injective.

To prove the converse implication (b)⇒(a), suppose $dF_{\overline{z}}$ is injective, we have (33) implies $(v_x; H) = 0$. On the one hand, fixing $v_x = 0$, (33) is equivalent to (35). Thus, the latter must imply $H = 0$, and hence by Lemma 2 (b), the W-SRCQ holds at $\overline{z}$. On the other hand, suppose $v_x \in \mathrm{appl}(\overline{z})$. Then the equation $\xi_{G(\overline{z})} H = \nabla g(\overline{x})^* v_x$ has a unique solution H, and it follows that for such (v_x, H), (33) is equivalent to (34). Therefore, the latter implies $v_x = 0$, which, by Lemma 2 (a) and $v_x \in \mathrm{appl}(\overline{z})$, is precisely the W-SOC at $\overline{z}$.

Next, we show that (b) $\iff$ (c) under the assumption that $\overline{z}$ is a KKT pair. For (b)⇒(c), applying [27, Proposition 5.18], there exist a neighborhood $\mathcal{U} \subseteq \mathcal{M}$ of $\overline{z}$ and a neighborhood $\mathcal{U}'$ containing $F(\overline{z})$, embedded in $\mathbb{X} \times \mathbb{S}^n$, such that F is a diffeomorphism from $\mathcal{U}$ onto $\mathcal{U}'$. Denote the restriction of F to $\mathcal{U}$ by $F_{\mathcal{U}}$. Consequently, $(F_{\mathcal{U}})^{-1}$ is single-valued and smooth on $\mathcal{U}'$, which yields (c).

For (c)⇒(b), we know (c) implies that $F_{\mathcal{U}}$ is bijective and $F_{\mathcal{U}}^{-1}$ is Lipschitz. Therefore, $d(F_{\mathcal{U}}^{-1})$ exists in an open dense subset $\mathcal{U}''$ in $\mathcal{U}'$. Hence $d(F_{\mathcal{U}}) \circ d(F_{\mathcal{U}}^{-1}) = id$ holds in $\mathcal{U}''$. Since $d(F_{\mathcal{U}})$ is always smooth and has below-bounded norm, $d(F_{\mathcal{U}}^{-1})$ can be smoothly extended to $\mathcal{U}'$. Therefore, $d(F_{\mathcal{U}})$ is always invertible in $\mathcal{U}$, which implies (b). □

By the implication *(a)⇒(c)* in Theorem 2, we immediately obtain the following stratum-restricted local error bound, which is the natural analogue of the classical local error bound in Definition 10.

Corollary 1 (The stratum-restricted local error bound) *For a KKT pair $\overline{z}$ of* (1)*, if the W-SOC and the W-SRCQ at $\overline{z}$, then there exists $C_{REB} > 0$ and a neighborhood $\mathcal{U} \subseteq \widetilde{\mathcal{M}}$ of $\overline{z}$ such that $F^{-1}_{\widetilde{\mathcal{M}}}(0) \cap \mathcal{U} = \{\overline{z}\}$ and*

$$\|F(z)\| \geq C_{REB} \|z - \overline{z}\| \quad \forall\, z \in \widetilde{\mathcal{M}} \cap \mathcal{U}. \tag{36}$$

We refer to (36) *as the stratum-restricted local error bound.*

Next, we present a simple semidefinite program to illustrate that the W-SOC and the W-SRCQ are strictly weaker than the problem-level regularity conditions commonly imposed in the existing literature. The same example also shows that, although the local error bound condition (Definition 10) fails, the *stratum-restricted local error bound* (36) still holds.

Example 1 Consider the following instance of (1):

$$\begin{aligned}
\min_{x \in \mathbb{R}^5} \quad & x_1 \\
\text{s.t.} \quad & \begin{bmatrix} 1 & 0 & 0 & x_4 + x_5 \\ 0 & x_4 - x_5 & 0 & x_3 \\ 0 & 0 & 0 & x_2 \\ x_4 + x_5 & x_3 & x_2 & x_1 \end{bmatrix} \in \mathbb{S}^4_+.
\end{aligned} \tag{37}$$

Clearly, $\overline{x} = [0; 0; 0; 0; 0]$ is an optimal solution, and $\overline{y} = \mathrm{Diag}(0, 0, 0, -1)$ is a corresponding Lagrange multiplier. Using Lemma 2, one can verify that both the W-SOC and W-SRCQ hold at $(\overline{x}, \overline{y})$, whereas the SOSC and SRCQ do not hold (and consequently, neither do the S-SOSC nor the constraint nondegeneracy). Therefore, by [19,

Proposition 5], $\mathcal{U}_0$ and $\mathcal{U}_{\mathcal{I}}$, the commonly used generalized Jacobians of F, are degenerate, which compromises the stability and fast convergence of standard semismooth Newton methods. Moreover, let

$$y(t) = \begin{bmatrix} 0 & 0 & 0 & t^2 \\ 0 & -t & 0 & 0 \\ 0 & 0 & 0 & 0 \\ t^2 & 0 & 0 & 1 \end{bmatrix}$$

with $t > 0$. A direct calculation yields $\|F(\overline{x}, y(t))\| = \Theta(t^2)$, but $\operatorname{dist}\big((\overline{x}, y(t)), F^{-1}(0)\big) = \Theta(t)$, where $\Theta(\cdot)$ denotes the tight asymptotic bound. Therefore, the local error bound condition does not hold at $(\overline{x}, \overline{y})$. On the other hand, let $\overline{z} = (\overline{x}, \overline{y})$. By (25), we know that for $v \in \mathcal{T}_{\overline{z}}\widetilde{\mathcal{M}}$,

$$dF_{\overline{z}}(v) = 0 \iff \begin{bmatrix} v_1 + H_{44} \\ 2v_2 + 2H_{34} \\ 2v_3 + 2H_{24} \\ 3v_4 + v_5 + H_{22} + 2H_{14} \\ v_4 + 3v_5 - H_{22} + 2H_{14} \end{bmatrix} = 0 \text{ and } \begin{bmatrix} H_{11} & H_{12} & H_{13} & v_4 + v_5 + \frac{1}{2}H_{14} \\ H_{21} & v_4 - v_5 & 0 & v_3 \\ H_{31} & 0 & 0 & v_2 \\ v_4 + v_5 + \frac{1}{2}H_{14} & v_3 & v_2 & v_1 \end{bmatrix} = 0,$$

where we have used the isomorphism from v to (v_x, H) in Lemma 1, with $v_x = (v_1; v_2; v_3; v_4; v_5)$ and $H = (H_{ij})_{4\times 4} \in \mathcal{T}_{G(\overline{z})}\mathcal{M}$. Using the characterization $\mathcal{T}_{G(\overline{z})}\mathcal{M} = \{H \in \mathbb{S}^4 \mid H_{22} = H_{23} = H_{33} = 0\}$, it is straightforward to verify that $dF_{\overline{z}}$ is injective. Together with [27, Proposition 5.18] and Corollary 1, this implies the validity of the stratum-restricted local error bound.

By continuity, the injectivity of dF_z at a given point persists in a neighborhood of that point. This observation enables us to establish the key properties needed in our analysis, namely the uniform boundedness of the inverse of the associated (almost) Hessian operator and the local uniqueness of the solution.

Corollary 2 *Let $\overline{z} = (\overline{x}, \overline{y}) \in \mathbb{X} \times \mathbb{S}^n$ with an IED $(\alpha, \beta, \gamma, p, q, \overline{P}, \overline{\lambda})$ of $G(\overline{z})$. If the W-SOC and the W-SRCQ holds at $\overline{z}$, then there exists a neighborhood $\mathcal{U} \subseteq \widetilde{\mathcal{M}}$ of $\overline{z}$ such that for any $z \in \mathcal{U}$, the following hold:*

(a) The manifold differential dF_z is injective;
(b) There exists a constant $C_{ub} > 0$ only depended on $\overline{z}$ that $\|(dF_z^* dF_z)^{-1}\| \le C_{ub}$;

Proof Firstly, since we always consider $z \in \widetilde{\mathcal{M}}$, we know the manifold at z is identical to the manifold $\mathcal{M}$ at $\overline{z}$ and the same holds for the index sets α, β and γ.

By Theorem 2, we know that $dF_{\overline{z}}$ is injective, which implies the invertibility of $dF_{\overline{z}}^* dF_{\overline{z}}$. Then, by continuity, there exists a neighborhood $\mathcal{U} \subseteq \widetilde{\mathcal{M}}$ of $\overline{z}$ such that for any $z \in \mathcal{U}$, dF_z remains injective and $\|(dF_z^* dF_z)^{-1}\|$ is uniformly bounded. □

3.1.2 Transversality, stability and genericity

We have established above that the W-SOC and the W-SRCQ are equivalent to the stratum-restricted strong metric regularity of the KKT mapping. In this subsection we further examine the geometric content of the W-SRCQ on a fixed stratum.

For problem (1), it is well-known that constraint nondegeneracy holds at a feasible point x if and only if the constraint mapping g intersects the manifold $\mathcal{M}_{p,0}$ transversally (see [6, (4.182)]). In the same spirit, we interpret the W-SRCQ (Definition 6) in terms of a transversality relation and investigate its stability and genericity within a prescribed stratum.

To place the W-SRCQ into the transversality framework (Definition 11), we first specify the geometric objects that will play the role of the target manifolds. For the subsequent stability analysis, it is convenient to work in a parametric setting and to introduce these objects in a neighborhood of the point of interest, rather than only at a single point. We begin by defining, for each $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$, a family of translated submanifolds

$$\mathcal{N}_z := \mathcal{M}_{p,0} + \Pi_{\mathbb{S}^n_-}\big(G(z)\big) - y, \tag{38}$$

where p is the positive inertia index of $G(z) = g(x) + y$ and $\mathbb{S}^n_-$ denotes the cone of negative semidefinite matrices. Constraint nondegeneracy (Definition 17) can be illustrated by $\mathcal{N}_z$ in the following proposition.

Proposition 2 *For the optimization problem (1), for any $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$ with an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$ of $G(z)$ and the corresponding manifolds $\mathcal{M}_{p,q}$, the following are equivalent:*

(i) $g \pitchfork_x \mathcal{N}_z$ in $\mathbb{S}^n$;

(ii) the constraint nondegeneracy holds (Definition 17) at $z = (x, y)$.

Proof Computing $\mathcal{T}_{g(x)}\mathcal{N}_z$ directly, we get

$$\begin{aligned}\mathcal{T}_{g(x)}\mathcal{N}_z &= \mathcal{T}_{G(z)}\big(\mathcal{M}_{p,0} + \Pi_{\mathbb{S}^n_-}(G(z))\big)\\ &= \big\{PBP^\top \mid B \in \mathbb{S}^n,\ B_{\beta\beta} = 0,\ B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0\big\}\end{aligned}$$

Therefore, $g \pitchfork_x \mathcal{N}_z$ if and only if

$$g'(x)\mathbb{X} + \big\{PBP^\top \mid B \in \mathbb{S}^n,\ B_{\beta\beta} = 0,\ B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0\big\} = \mathbb{S}^n,$$

which is exactly Definition 17. □

Remark 6 When $z = (x, y) \in \mathbb{X}\times\mathbb{S}^n$ is a KKT pair of (1), one can verify that $\mathcal{N}_z = \mathcal{M}_{p,0}$, and the transversality condition $g \pitchfork_x \mathcal{M}_{p,0}$ in $\mathbb{S}^n$ coincides precisely with the classical constraint nondegeneracy condition (Definition 3) given in [6, Definition 5.70].

The next lemma shows that, for $z \in \widetilde{\mathcal{M}}_{p,q}$, the set $\mathcal{N}_z$ is locally embedded in the stratum $\mathcal{M}_{p,q}$ around $G(z)$, in the sense made precise below. The proof is deferred to Appendix B.

Lemma 3 *Let $z \in \widetilde{\mathcal{M}}_{p,q}$ and set $A := G(z) \in \mathcal{M}_{p,q}$. Write*

$$A_+ := \Pi_{\mathbb{S}^n_+}(A) \in \mathcal{M}_{p,0}, \quad A_- := \Pi_{\mathbb{S}^n_-}(A) \in \mathcal{M}_{0,q},$$

so that $A = A_+ + A_-$. Then there exists a neighborhood $\mathcal{U}$ of A in $\mathcal{M}_{p,0} + A_-$ such that $\mathcal{U}$ is smoothly embedded in $\mathcal{M}_{p,q}$. Therefore, there exists a neighborhood $\mathcal{U}$ of $g(x) = A - y$ in $\mathcal{M}_{p,0} + A_- - y$ such that $\mathcal{U}$ is smoothly embedded in $\mathcal{M}_{p,q} - y$, which by definition means $\mathcal{N}_z$ is locally embedded in $\mathcal{M}_{p,q} - y$ near $g(x)$.

We now enlarge $\mathcal{N}_z$ for $z = (x, y) \in \widetilde{\mathcal{M}}_{p,q}$ by taking into account the normal directions to the stratum. For each $A \in \mathcal{U} \subseteq \mathcal{N}_z \cap (\mathcal{M}_{p,q} - y)$, consider the normal space of $\mathcal{M}_{p,q} - y$ in $\mathbb{S}^n$ at A, which admits the representation

$$\mathcal{N}_A(\mathcal{M}_{p,q} - y) = \big\{PHP^\top \mid \text{only the block } H_{\beta\beta} \text{ may be nonzero}\big\},$$

where $(\alpha, \beta, \gamma, p, q, P, \lambda)$ is an IED of $A + y$. We then form the union

$$\mathcal{DN}_z := \bigcup_{A\in\mathcal{N}_z} \big(A + \mathcal{N}_A(\mathcal{M}_{p,q} - y)\big) = \big\{A + W \mid A \in \mathcal{N}_z,\ W \in \mathcal{N}_A(\mathcal{M}_{p,q} - y)\big\} \subseteq \mathbb{S}^n. \tag{39}$$

The following theorem is the main result of this part and will relate the W-SRCQ to a suitable transversality condition involving the family $\{\mathcal{DN}_z\}$.

Theorem 3 *For the optimization problem (1), for any $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$ with an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$ of $G(z)$ and the corresponding manifolds $\mathcal{M}_{p,q}$, the following are equivalent:*

(i) $g \pitchfork_x \mathcal{DN}_z$ in $\mathbb{S}^n$;
(ii) the W-SRCQ holds (Definition 15) at $z = (x, y)$.

The rigorous proof is deferred to Appendix B. Here we briefly explain the main ideas. First, although there are notions of transversality formulated for general sets, in this paper, we work with the classical transversality theory for smooth manifolds. Consequently, we need to verify that the sets $\mathcal{DN}_z$ are indeed smooth manifolds (around the points of interest) for all relevant z. Establishing this manifold structure is the main technical step, and the details are given in the appendix. Next, we explain informally why the transversality condition in the theorem reduces to the W-SRCQ, for readers who prefer to avoid differential-topological arguments. By construction, $\mathcal{DN}_z$ is obtained by attaching the normal space $\mathcal{N}_A\mathcal{M}_{p,q}$ to each point $A \in \mathcal{N}_z$ (and then shifting by $-y$). Thus, at the point $g(x) = G(z) - y$, we have

$$\mathcal{T}_{g(x)}\mathcal{DN}_z = \mathcal{T}_{G(z)}(\mathcal{DN}_z + y) = \mathcal{T}_{G(z)}\mathcal{N}_z \oplus \mathcal{N}_{G(z)}\mathcal{M}_{p,q} = \big\{PBP^\top \;\big|\; B \in \mathbb{S}^n,\ B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0\big\}.$$

Therefore, the transversality relation $g \pitchfork_x \mathcal{DN}_z$ in $\mathbb{S}^n$ can be translated to

$$dg_x(\mathbb{X}) + \big\{PBP^\top \;\big|\; B \in \mathbb{S}^n,\ B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0\big\} = \mathbb{S}^n, \tag{40}$$

which is precisely the W-SRCQ condition.

These concepts may seem abstract at first, but they become more intuitive when viewed through a geometric lens, as illustrated in the following figure.

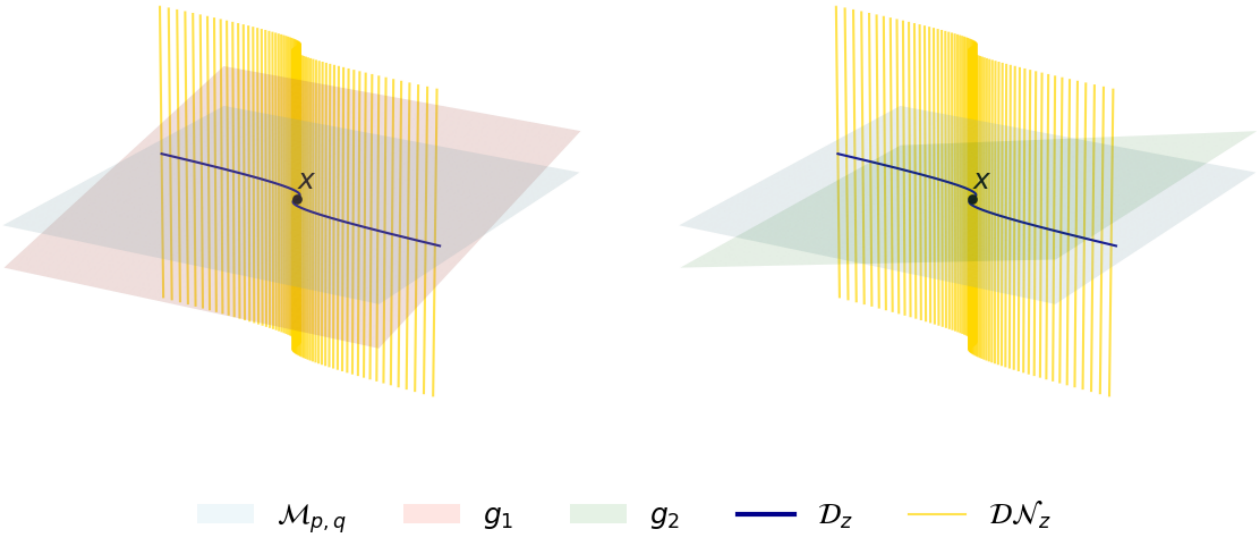

Fig. 2: The visualization of transversality

Figure 2 gives an imprecise but understandable visual illustration of the transversality characterization of the W-SRCQ and constraint nondegeneracy. Let $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$. The ambient space represents $\mathbb{S}^n$. The light blue plane represents $\mathcal{M}_{p,q} - y$ embedded in $\mathbb{S}^n$, and the dark blue line represents $\mathcal{N}_z \subseteq \mathcal{M}_{p,q} - y$ (see (38)). The yellow bundle illustrates $\mathcal{DN}_z$ (see (39)), which is the normal bundle of $\mathcal{M}_{p,q} - y$ restricted to $\mathcal{N}_z$. The red plane and the green plane represent two maps g_1 and g_2, which can be interpreted as the images of two embedding maps, respectively. In the left figure, $g_1 \pitchfork_x \mathcal{N}_z$, implying that both constraint nondegeneracy and the W-SRCQ hold at the point z. In the right figure, $g_2 \pitchfork_x \mathcal{DN}_z$ but g_2 does not transverse $\mathcal{N}_z$, meaning that the W-SRCQ holds while constraint nondegeneracy fails at z.

The following corollary reveals some relations between the constraint nondegeneracy (Definition 17) and the W-SRCQ (Definition 15).

Corollary 3 *For $z = (x, y) \in \widetilde{\mathcal{M}}_{p,q}$ such that $p + q = n$, then W-SRCQ holds at z if and only if constraint nondegeneracy (Definition 17) holds at z.*

Proof Note that when $p + q = n$, $\mathcal{M}_{p,q}$ will be an open submanifold of $\mathbb{S}^n$. So

$$\mathcal{N}_{g(x)}(\mathcal{M}_{p,q} - y) = \mathcal{N}_{G(z)}\mathcal{M}_{p,q} = \{0\}.$$

Then $\mathcal{DN}_z = \mathcal{N}_z$. Then by Proposition 2 and Theorem 3, we get what we want. □

This corollary shows that constraint nondegeneracy can be viewed as a special case of the W-SRCQ. Nevertheless, constraint nondegeneracy occupies a distinguished position in the theory. In the next subsection we shall see that it plays a particularly important role when one investigates variational properties *across* different strata.

On the other hand, the classical transversality characterization of constraint nondegeneracy [6, (4.182)] implies that constraint nondegeneracy is stable under small perturbations. This naturally raises the question of whether an analogous stability property holds for the W-SRCQ. The following theorem answers this question in the affirmative and establishes the stability of the W-SRCQ along the stratum $\widetilde{\mathcal{M}}_{p,q}$ via its transversality characterization.

Theorem 4 *If W-SRCQ holds at some $\overline{z} = (\overline{x}, \overline{y}) \in \widetilde{\mathcal{M}}_{p,q}$, then there will exist a neighborhood $\widetilde{\mathcal{U}} \subseteq \widetilde{\mathcal{M}}_{p,q}$ of $\overline{z}$ such that W-SRCQ holds for any $z \in \widetilde{\mathcal{U}}$.*

The rigorous proof is deferred to Appendix B. Here we briefly indicate the underlying reason. Transversality is, in essence, a first-order nonsingularity condition: it requires that the sum of the relevant tangent spaces spans the ambient space. Such a full-rank condition is open, and therefore it persists under sufficiently small perturbations of the data. Consequently, once the W-SRCQ is characterized by an appropriate transversality relation, its stability follows immediately from the openness of transversality.

Remark 7 A more elementary argument will be given in the next subsection. While stability along a fixed stratum has already been established in Proposition 3, our purpose here is to emphasize the geometric nature of the W-SRCQ and to highlight its close parallel with the classical constraint nondegeneracy condition.

To conclude this subsection, we turn to another fundamental feature of the W-SRCQ, namely its *genericity*, which parallels the classical situation for constraint nondegeneracy. Fix $p, q \in \mathbb{Z}_+$ and let $g : \mathbb{X} \to \mathbb{S}^n$ be a smooth mapping. For a perturbation parameter $b \in \mathbb{S}^n$, define the perturbed mapping

$$g_b(x) := g(x) + b, \qquad x \in \mathbb{X},$$

and, for $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$, set the associated shifted quantity accordingly.

Theorem 5 *For generic $b \in \mathbb{S}^n$, W-SRCQ holds at all $z = (x, y)$ such that $\mathbb{S}^n_+ \ni g_b(x) \perp y \in \mathbb{S}^n_-$.*

As before, the rigorous proof is deferred to Appendix B.

Remark 8 Having established the geometric interpretation of the W-SRCQ via transversality, we can naturally derive that of the W-SOC through the dual characterization provided in [19, Section 4], assuming the primal problem (1) is a convex quadratic program. However, obtaining a primal geometric characterization of the W-SOC and the S-SOSC is challenging. Although a more complicated version could be developed, we find it would be far from practical use. Hence, we leave it as a direction for future work.

3.2 Variational properties across strata

Building on the results of the previous subsection, we now have a detailed picture of the variational behavior *within* each stratum of the index stratification. The purpose of the present subsection is to investigate how these properties interact *across* different strata. We will show that the newly introduced problem-level regularity conditions, as well as the solution-level regularity conditions, exhibit natural openness and closedness properties when restricted to strata, with the precise behavior governed by the adjacency relations among strata. Furthermore, we prove that the classical strong-form problem-level regularity conditions are equivalent to the local *uniform validity* of their weak-form counterparts.

As we are discussing perturbation properties, one should keep in mind that, unless explicitly stated otherwise, we will use the notation from Definitions 15–16, which have been extended to non-KKT pairs.

3.2.1 The W-SRCQ across strata

We begin by investigating perturbation properties of the W-SRCQ across strata. Our first result reveals a certain openness property of the W-SRCQ with respect to the index stratification: the W-SRCQ is *stable* under small perturbations, as long as the negative inertia index remains unchanged.

Proposition 3 *Let $\overline{z} = (\overline{x}, \overline{y}) \in \mathbb{X} \times \mathbb{S}^n$. Denote the negative inertia index of $G(\overline{z})$ by q. If the W-SRCQ holds at $\overline{z}$, then there exists a neighborhood $\mathcal{U}$ of $\overline{z}$ such that the W-SRCQ holds at all $z \in \mathcal{U}$ for which the negative inertia index of $G(z)$ is q.*

Proof Denote an IED of $G(\overline{z})$ as $(\alpha, \beta, \gamma, p, q, P, \lambda)$. Suppose the statement does not hold. Then there exists a sequence $z^\nu = (x^\nu, y^\nu) \in \mathbb{X} \times \mathbb{S}^n$ such that $z^\nu \to \overline{z}$, the negative inertia index of $G(z^\nu)$ is q, and the W-SRCQ fails at each z^ν. Denote an IED of $G(z^\nu)$ by $(\alpha^\nu, \beta^\nu, \gamma, p^\nu, q, P^\nu, \lambda^\nu)$. Since $z^\nu \to \overline{z}$, we have $G(z^\nu) \to G(\overline{z})$, and thus $\mathrm{dist}(P^\nu, \mathcal{O}^n(G(\overline{z}))) \to 0$ (see, e.g., [3, Theorem VII.3.1]). By passing to a subsequence if necessary, we may assume $P^\nu \to P'$ for some $P' \in \mathcal{O}^n(G(\overline{z}))$, and $\beta^\nu \equiv \beta'$ for some fixed index set β'. Note that due to continuity and eigenvalue convergence, we must have $\beta' \subseteq \beta$.

Now, by (31), the failure of W-SRCQ at z^ν implies that

$$g'(x^\nu)\mathbb{X} + \left\{P^\nu B (P^\nu)^\top \mid B \in \mathbb{S}^n,\ B_{\beta'\gamma} = 0,\ B_{\gamma\gamma} = 0\right\} \neq \mathbb{Y}.$$

Taking the limit as $\nu \to \infty$, and using the continuity of g' and the convergence $P^\nu \to P'$, we obtain

$$g'(\overline{x})\mathbb{X} + \left\{P' B (P')^\top \mid B \in \mathbb{S}^n,\ B_{\beta'\gamma} = 0,\ B_{\gamma\gamma} = 0\right\} \neq \mathbb{Y},$$

which contradicts the assumption that the W-SRCQ holds at $\overline{z}$, since $P' \in \mathcal{O}^n(G(\overline{z}))$ and $\beta' \subseteq \beta$. The proof is then completed. □

This proposition shows that, locally, the validity of the W-SRCQ propagates to higher-dimensional strata as long as the negative inertia index remains fixed, a behavior we interpret as a form of openness. By contrast, an analogous openness property for constraint nondegeneracy (Definition 17) holds without imposing any restriction on the negative inertia index.

Corollary 4 *Let $\overline{z} = (\overline{x}, \overline{y}) \in \mathbb{X} \times \mathbb{Y}$. If the constraint nondegeneracy holds at $\overline{z}$, then there exists a neighborhood $\mathcal{U}$ of $\overline{z}$ such that the constraint nondegeneracy holds at all $z \in \mathcal{U}$.*

Proof Denote the positive inertia index of $G(\overline{z})$ as $\overline{p}$ and define $\overline{\alpha} = [1, 2, \cdots, \overline{p}]$. Suppose the statement does not hold, then there exists a sequence $z^\nu = (x^\nu, y^\nu) \in \mathbb{X} \times \mathbb{Y}$ such that $z^\nu \to \overline{z}$ and constraint nondegeneracy fails at each z^ν. Denote an IED of $G(z^\nu)$ by $(\alpha^\nu, \beta^\nu, \gamma^\nu, p^\nu, q^\nu, P^\nu, \lambda^\nu)$. Since $z^\nu \to \overline{z}$, we have $G(z^\nu) \to G(\overline{z})$, and thus $\mathrm{dist}(P^\nu, \mathcal{O}^n(G(\overline{z}))) \to 0$ (see, e.g., [3, Theorem VII.3.1]). By passing to a subsequence if necessary, we may assume $P^\nu \to P$ for some $P \in \mathcal{O}^n(G(\overline{z}))$, and $\alpha^\nu \equiv \alpha$ for some fixed index set α. Note that due to continuity and eigenvalue convergence, we must have $\alpha \supseteq \overline{\alpha}$.

Now, by (32), the failure of constraint nondegeneracy at z^ν implies that

$$g'(x^\nu)\mathbb{X} + \left\{P^\nu B (P^\nu)^\top \mid B \in \mathbb{S}^n,\ B_{\beta\beta} = 0,\ B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0\right\} \neq \mathbb{Y}.$$

Taking the limit as $\nu \to \infty$, and using the continuity of g' and the convergence $P^\nu \to P$, we obtain

$$g'(\overline{x})\mathbb{X} + \left\{PBP^\top \mid B \in \mathbb{S}^n,\ B_{\beta\beta} = 0,\ B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0\right\} \neq \mathbb{Y}.$$

which contradicts the assumption that the constraint nondegeneracy holds at $\overline{z}$, since $P \in \mathcal{O}^n(G(\overline{z}))$ and $\alpha \supseteq \overline{\alpha}$. The proof is then completed. □

The following result is an immediate consequence of Corollary 4. It shows that constraint nondegeneracy holds if and only if the W-SRCQ holds collectively for all relevant strata in a neighborhood of the point under consideration.

Theorem 6 *Let $\overline{z} = (\overline{x}, \overline{y}) \in \mathbb{X} \times \mathbb{Y}$. Then the following three statements are equivalent:*

(a) The constraint nondegeneracy holds at $(\overline{x}, \overline{y})$;
(b) The W-SRCQ holds at every pair (x, y) with (x, y) sufficiently close to $(\overline{x}, \overline{y})$;
(c) The W-SRCQ holds at every pair $(\overline{x}, y)$ with y sufficiently close to $\overline{y}$.

Proof To prove (a)⇒(b), by Corollary 4, we know that (32) holds for (x, y) sufficiently close to $(\overline{x}, \overline{y})$, which then implies the validity of (31), i.e., the W-SRCQ. Next, (b)⇒(c) is immediate.

To prove (c)⇒(a), let $(\alpha, \beta, \gamma, p, q, P, \lambda)$ be an IED of $G(\overline{z})$, and define

$$y^t = \overline{y} - P \begin{bmatrix} 0 & 0 & 0 \\ 0 & tI & 0 \\ 0 & 0 & 0 \end{bmatrix} P^T,$$

where we choose $t > 0$ sufficiently small so that y^t is close to $\overline{y}$. Consequently, the W-SRCQ holds at $(\overline{x}, y^t)$, which, by definition, implies that the constraint nondegeneracy holds at $(\overline{x}, \overline{y})$. □

As the classical constraint nondegeneracy (Definition 3) can be viewed as a special case of the constraint nondegeneracy (Definition 17) studied in this section, the above lemma also provides a characterization of the classical constraint nondegeneracy. The proof is omitted, as it follows directly from the proof of Theorem 6.

Corollary 5 *Let $\overline{x}$ be a feasible point of (1) and $\overline{y} \in M(\overline{x})$. Then the following three statements are equivalent:*

(a) The constraint nondegeneracy (Definition 3) holds at $\overline{x}$;
(b) The W-SRCQ holds at every pair (x, y) with (x, y) sufficiently close to $(\overline{x}, \overline{y})$;
(c) There exists y strictly complementary to $g(\overline{x})$, such that the W-SRCQ holds at $(\overline{x}, y)$;

As noted above, Proposition 3 establishes a certain openness property of the W-SRCQ. The next result reveals a complementary closedness property of the W-SRCQ with respect to the index stratification. Roughly speaking, the validity of the W-SRCQ on a neighborhood within a stratum can be extended to certain parts of its closure, provided that the positive inertia index is preserved.

Proposition 4 *Let $\mathcal{U}$ be an open set in $\mathbb{X} \times \mathbb{Y}$ and let $\widetilde{\mathcal{M}}_{p,q}$ be a stratum of $\mathbb{S}^n$. If the W-SRCQ holds at every $z \in \mathcal{U} \cap \widetilde{\mathcal{M}}_{p,q}$ for some p, q, then the W-SRCQ holds at every $z \in \mathcal{U} \cap \widetilde{\mathcal{M}}_{p,q'}$ for all q' such that $0 \leq q' \leq q$.*

Proof For any q' such that $0 \le q' \le q$ and $z = (x, y) \in \mathcal{U} \cap \widetilde{\mathcal{M}}_{p,q'}$ with IED of $G(z)$ given by $(\alpha, \beta', \gamma', p, q', P, \lambda)$, define the matrix

$$\Delta = P \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -I & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} P^T,$$

where the block partition corresponds to $[p, n-p-q, q-q', q']$ and I is the identity matrix of order $q-q'$. Then the sequence $z^t = (x, y + t\Delta)$ lies in the stratum $\widetilde{\mathcal{M}}_{p,q}$ for all $t > 0$, and $z^t \to z$ as $t \downarrow 0$. Since $\mathcal{U}$ is open and contains z, there exists $t_0 > 0$ such that $z^t \in \mathcal{U}$ for all $t \in (0, t_0]$. By assumption, the W-SRCQ holds at each z^t for $t \in (0, t_0]$.

By (31), the W-SRCQ at z^t is equivalent to

$$g'(x)\mathbb{X} + \left\{ PBP^\top \mid B \in \mathbb{S}^n,\ B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0 \right\} = \mathbb{Y}, \tag{41}$$

where $\beta = [p+1, p+2, \ldots, n-q]$ and $\gamma = [n-q+1, n-q+2, \ldots, n]$. Because $\beta \subseteq \beta'$ and $\gamma \supseteq \gamma'$, it follows that the subspace in the W-SRCQ at z,

$$\left\{ PBP^\top \mid B \in \mathbb{S}^n,\ B_{\beta'\gamma'} = 0,\ B_{\gamma'\gamma'} = 0 \right\},$$

is contained in the corresponding subspace in (41). Therefore, the W-SRCQ holds at z. To conclude, the W-SRCQ holds at every $z \in \mathcal{U} \cap \widetilde{\mathcal{M}}_{p,q'}$ for all q' satisfying $0 \le q' \le q$. □

3.2.2 The W-SOC across strata

A dual counterpart of the results in Section 3.2.1 can be developed by exploiting the primal–dual relationships between the W-SOC and the W-SRCQ established in [19, Section 4]. In particular, for convex quadratic instances of (1), these relationships yield dual versions of the stratified perturbation statements proved earlier. For general (possibly nonconvex) problems (1), the same conclusions can be obtained similarly to those of Section 3.2.1 by the appropriate dual arguments. For instance, the following proposition establishes a stratified openness property of the W-SOC. It should be viewed as the dual analogue of Proposition 3. The key difference is that the *positive* inertia index p is kept fixed here (rather than the negative inertia index q).

Proposition 5 *Let $\overline{z} = (\overline{x}, \overline{y}) \in \mathbb{X} \times \mathbb{Y}$, and let p be the positive inertia index of $G(\overline{z})$. Suppose that the W-SOC holds at $\overline{z}$. Then there exists a neighborhood $\mathcal{U}$ of $\overline{z}$ such that the W-SOC holds at every $z \in \mathcal{U}$ for which $G(z)$ has positive inertia index p.*

Proof Denote an IED of $G(\overline{z})$ by $(\alpha, \beta, \gamma, p, q, P, \lambda)$. Suppose, for contradiction, that the statement does not hold. Then there exists a sequence $z^\nu = (x^\nu, y^\nu) \in \mathbb{X} \times \mathbb{Y}$ such that $z^\nu \to \overline{z}$, the positive inertia index of $G(z^\nu)$ is p, and the W-SOC fails at each z^ν. Let an IED of $G(z^\nu)$ be denoted by $(\alpha, \beta^\nu, \gamma^\nu, p, q^\nu, P^\nu, \lambda^\nu)$. Since $z^\nu \to \overline{z}$, we have $G(z^\nu) \to G(\overline{z})$, and thus $\operatorname{dist}(P^\nu, \mathcal{O}^n(G(\overline{z}))) \to 0$ (see, e.g., [3, Theorem VII.3.1]). By passing to a subsequence if necessary, we may assume $P^\nu \to P'$ for some $P' \in \mathcal{O}^n(G(\overline{z}))$, and $(\beta^\nu, \gamma^\nu) \equiv (\beta', \gamma')$ for some fixed index sets β', γ'. Note that, by continuity and eigenvalue convergence, $\gamma' \supseteq \gamma$.

Now, by Definition 14, the failure of W-SOC at z^ν implies the existence of $d^\nu \in \mathbb{X}$ with $\|d^\nu\| = 1$ such that

$$\begin{aligned} &\left\langle d^\nu, \nabla^2_{xx} L(z^\nu) d^\nu \right\rangle + 2 \sum_{i \in \alpha} \sum_{j \in \gamma'} \frac{-\lambda^\nu_j}{\lambda^\nu_i} \left[(P^\nu)^\top (\nabla g(x^\nu)^* d^\nu) P^\nu \right]^2_{ij} = 0, \\ &[(P^\nu)^\top (\nabla g(x^\nu)^* d^\nu) P^\nu]_{\beta'\beta'} = 0, \quad [(P^\nu)^\top (\nabla g(x^\nu)^* d^\nu) P^\nu]_{\beta'\gamma'} = 0, \quad \text{and} \quad [(P^\nu)^\top (\nabla g(x^\nu)^* d^\nu) P^\nu]_{\gamma'\gamma'} = 0. \end{aligned}$$

By taking a further subsequence if necessary, assume $d^\nu \to d^\infty \in \mathbb{X}$ where $\|d^\infty\| = 1$. Taking the limit as $\nu \to \infty$ and using the continuity of the involved functions and eigenvectors, we obtain

$$\begin{aligned} &\left\langle d^\infty, \nabla^2_{xx} L(\overline{z}) d^\infty \right\rangle + 2 \sum_{i \in \alpha} \sum_{j \in \gamma'} \frac{-\overline{\lambda}_j}{\overline{\lambda}_i} \left[(P')^\top (\nabla g(\overline{x})^* d^\infty) P' \right]^2_{ij} = 0, \\ &[(P')^\top (\nabla g(\overline{x})^* d^\infty) P']_{\beta'\beta'} = 0, \quad [(P')^\top (\nabla g(\overline{x})^* d^\infty) P']_{\beta'\gamma'} = 0, \quad \text{and} \quad [(P')^\top (\nabla g(\overline{x})^* d^\infty) P']_{\gamma'\gamma'} = 0. \end{aligned}$$

This contradicts the assumption that W-SOC holds at $\overline{z}$, since $P' \in \mathcal{O}^n(G(\overline{z}))$, $\gamma' \supseteq \gamma$, $\beta' \cup \gamma' = \alpha^c$, and $\overline{\lambda}_j = 0$ for all $j \in \gamma' \setminus \gamma$. The proof is therefore complete. □

Recall from Section 3.2.1 that the openness property of the W-SRCQ (Proposition 5) and the constraint nondegeneracy (Proposition 4) lead to Theorem 6, which characterizes constraint nondegeneracy via the local collective validity of the W-SRCQ. For second order conditions, however, the picture is more subtle: the strong second order sufficient condition (S-SOSC) is not, in general, stable in the same stratified sense. Consequently, establishing a dual analogue of Theorem 6, that is, relating the S-SOSC to the local collective validity of the W-SOC, requires additional assumptions and more delicate technical tools. As a first step in this direction, we introduce a second order necessary condition (SONC) at a primal-dual pair.

Definition 18 Let $z = (x, y) \in \mathbb{X} \times \mathbb{Y}$, and let $(\alpha, \beta, \gamma, p, q, P, \lambda)$ be an IED of $G(z)$. We say that the *second order necessary condition (SONC)* holds at z if

$$\left\langle d, \nabla^2_{xx} L(x, y)\, d \right\rangle - 2 \sum_{i \in \alpha} \sum_{j \in \gamma} \frac{\lambda_j}{\lambda_i} \Big(\big[P^\top (\nabla g(x)^* d) P \big]_{ij} \Big)^2 \geq 0 \tag{42}$$

for every nonzero $d \in \mathbb{X}$ satisfying

$$\big[P^\top (\nabla g(x)^* d) P \big]_{\beta\beta} \succeq 0, \quad \big[P^\top (\nabla g(x)^* d) P \big]_{\beta\gamma} = 0 \quad \text{and} \quad \big[P^\top (\nabla g(x)^* d) P \big]_{\gamma\gamma} = 0. \tag{43}$$

Remark 9 The classical SONC (see, for example, [5, Theorem 3.1 and (3.20)]) is defined at a stationary point $\overline{x}$, and involves a supremum over the set of Lagrange multipliers $M(\overline{x})$ in the quadratic form. In contrast, Definition 18 is defined at an arbitrary primal-dual pair $z = (x, y)$, which is natural in our setting since the analysis is carried out on the full space $\mathbb{X} \times \mathbb{Y}$. Moreover, when $\overline{x}$ is stationary and $M(\overline{x})$ is a singleton, Definition 18 reduces to the classical SONC.

Observe that if a self-adjoint linear operator $\mathcal{Q}$ on a Euclidean space $\mathbb{X}$ satisfies

$$\langle \mathcal{Q}x, x \rangle \neq 0 \quad \forall\, x \in \mathcal{H} \setminus \{0\},$$

for some subspace $\mathcal{H} \subseteq \mathbb{X}$, then by continuity of the mapping $x \mapsto \langle \mathcal{Q}x, x \rangle$ and compactness of the unit sphere in $\mathcal{H}$, there exists a constant $c > 0$ such that

$$|\langle \mathcal{Q}x, x \rangle| \geq c \|x\|^2 \quad \forall\, x \in \mathcal{H}.$$

Motivated by this elementary fact, we introduce a uniform version of the W-SOC in a neighborhood of a given point $\overline{z} \in \mathbb{X} \times \mathbb{S}^n$.

Definition 19 Consider problem (1). Let $\overline{z} = (\overline{x}, \overline{y}) \in \mathbb{X} \times \mathbb{S}^n$, and $\mathcal{U}$ be a set containing $\overline{z}$. We say that the *weak second order condition (W-SOC)* holds *uniformly on* $\mathcal{U}$ if there exists a constant $c > 0$ such that, for every $z = (x, y) \in \mathcal{U}$ with an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$ of $G(z)$, one has

$$\left\langle v_x, \nabla^2_{xx} L(x, y)\, v_x \right\rangle + 2 \sum_{i \in \alpha} \sum_{j \in \gamma} \frac{-\lambda_j}{\lambda_i} \Big(\big[P^\top (\nabla g(x)^* v_x) P \big]_{ij} \Big)^2 \geq c \|v_x\|^2 \quad \forall\, v_x \in \operatorname{appl}(x, y), \tag{44}$$

where $\operatorname{appl}(x, y)$ is defined in (30).

With Definitions 18 and 19 in place, we can now establish the desired relationship between the S-SOSC and the uniform validity of the W-SOC.

Theorem 7 *Let $\overline{z} = (\overline{x}, \overline{y}) \in \mathbb{X} \times \mathbb{Y}$. Consider the following statements:*

(a) The S-SOSC holds at $\overline{z}$.
(b) The W-SOC holds uniformly on a neighborhood of $\overline{z}$.
(c) The W-SOC holds uniformly for $(\overline{x}, y)$ with y in a neighborhood of $\overline{y}$.

Then (a)⇒(b)⇒(c)*. Moreover, if the SONC holds at $\overline{z}$, then* (a), (b)*, and* (c) *are equivalent.*

Proof Denote an IED of $G(\overline{z})$ by $(\overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{p}, \overline{q}, \overline{P}, \overline{\lambda})$.

To prove (a)⇒(b), we proceed by contradiction. Suppose there exists a sequence $z^\nu = (x^\nu, y^\nu) \to \overline{z}$ and $c^\nu \downarrow 0$ such that the W-SOC with constant c^ν fails at each z^ν. Let an IED of $G(z^\nu)$ be $(\alpha^\nu, \beta^\nu, \gamma^\nu, p^\nu, q^\nu, P^\nu, \lambda^\nu)$. Since $z^\nu \to \overline{z}$, by passing to a subsequence if necessary, we may assume $P^\nu \to P \in \mathcal{O}^n(G(\overline{z}))$, $(\alpha^\nu, \beta^\nu, \gamma^\nu, p^\nu, q^\nu) \equiv (\alpha, \beta, \gamma, p, q)$, and $\lambda^\nu \to \overline{\lambda}$. Without loss of generality, assume $\alpha \supseteq \overline{\alpha}$, $\beta \subseteq \overline{\beta}$, $\gamma \supseteq \overline{\gamma}$.

By (44), since the W-SOC with constant c^ν fails at z^ν, there exists $d^\nu \in \mathbb{X}$ with $\|d^\nu\| = 1$ such that

$$\left| \left\langle d^\nu, \nabla^2_{xx} L(z^\nu) d^\nu \right\rangle + 2 \sum_{i \in \alpha} \sum_{j \in \gamma} \frac{-\lambda^\nu_j}{\lambda^\nu_i} \big[(P^\nu)^\top (\nabla g(x^\nu)^* d^\nu) P^\nu \big]^2_{ij} \right| < c^\nu \tag{45}$$

and $d^\nu \in \mathrm{appl}(z^\nu)$, i.e.,

$$[(P^\nu)^\top(\nabla g(x^\nu)^* d^\nu)P^\nu]_{\beta\beta} = 0,\ [(P^\nu)^\top(\nabla g(x^\nu)^* d^\nu)P^\nu]_{\beta\gamma} = 0,\ [(P^\nu)^\top(\nabla g(x^\nu)^* d^\nu)P^\nu]_{\gamma\gamma} = 0. \tag{46}$$

For $i \in \alpha \setminus \overline{\alpha}$ and $j \in \overline{\gamma}$, we have $\lambda_i^\nu \downarrow 0$ and $\lambda_j^\nu \to \overline{\lambda}_j < 0$, so $\frac{-\lambda_j^\nu}{\lambda_i^\nu} \to +\infty$. Since $\left\{\langle d^\nu, \nabla^2_{xx} L(z^\nu) d^\nu\rangle\right\}_\nu$ is bounded, combining (45), we must have

$$\left[(P^\nu)^\top(\nabla g(x^\nu)^* d^\nu)P^\nu\right]_{ij} \to 0 \quad \forall\, i \in \alpha \setminus \overline{\alpha},\ j \in \overline{\gamma}. \tag{47}$$

Moreover, we have

$$\left\langle d^\nu, \nabla^2_{xx} L(z^\nu) d^\nu \right\rangle + 2 \sum_{i\in\overline{\alpha}} \sum_{j\in\overline{\gamma}} \frac{-\lambda_j^\nu}{\lambda_i^\nu} \left[(P^\nu)^\top(\nabla g(x^\nu)^* d^\nu)P^\nu\right]^2_{ij} < c^\nu. \tag{48}$$

Without loss of generality, we may assume that $d^\nu \to d^\infty$ for some $d^\infty \in \mathbb{X}$ and $\|d^\infty\| = 1$. Then, by passing ν to ∞ in (48), (46), and (47), we get

$$\begin{aligned}
&\left\langle d^\infty, \nabla^2_{xx} L(\overline{z}) d^\infty \right\rangle + 2 \sum_{i\in\overline{\alpha}} \sum_{j\in\overline{\gamma}} \frac{-\overline{\lambda}_j}{\overline{\lambda}_i} \left[[P^\top(\nabla g(\overline{x})^* d^\infty)P]\right]^2_{ij} \le 0,\\
&[P^\top(\nabla g(\overline{x})^* d^\infty)P]_{\beta\beta} = 0,\ [P^\top(\nabla g(\overline{x})^* d^\infty)P]_{\beta\gamma} = 0,\ [(P^\top(\nabla g(\overline{x})^* d^\infty)P]_{\gamma\gamma} = 0,\\
&\text{and} \quad [(P^\top(\nabla g(\overline{x})^* d^\infty)P]_{ij} = 0 \quad \forall\, i \in \alpha \setminus \overline{\alpha},\ j \in \overline{\gamma}
\end{aligned}$$

which together shows that we find a nonzero $d^\infty \in \mathrm{appl}(\overline{z})$ but contradicts the S-SOSC at $\overline{z}$. Thus, we have (a)⇒(b). Clearly, (b)⇒(c) is immediate.

Next, we show (c)⇒(a) under the assumption of the SONC. Define the matrix

$$\Delta = \overline{P} \begin{bmatrix} 0 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{bmatrix} \overline{P}^T,$$

where the block partition corresponds to $[\overline{p}, n-\overline{p}-\overline{q}, \overline{q}]$ and I is the identity matrix of order $n-\overline{p}-\overline{q}$. Then the sequence $z^t = (\overline{x}, \overline{y} + t\Delta) \to \overline{z}$ as $t \downarrow 0$. By assumption (c), there is $t_0 > 0$ such that the W-SOC holds uniformly at z^t for $t \in (0, t_0]$.

By (44), the uniform validity of the W-SOC shows that there is c such that, for any $t \in (0, t_0]$,

$$\left\langle v_x, \nabla^2_{xx} L(z^t) v_x \right\rangle + 2 \sum_{i\in\overline{\gamma}} \sum_{j\in\overline{\gamma}} \frac{-\overline{\lambda}_j}{\overline{\lambda}_i} \left[\overline{P}^\top(\nabla g(\overline{x})^* v_x)\overline{P}\right]^2_{ij} + 2 \sum_{\overline{p}+1\le i\le n-\overline{q}} \sum_{j\in\overline{\gamma}} \frac{-\overline{\lambda}_j}{t} \left[\overline{P}^\top(\nabla g(\overline{x})^* v_x)\overline{P}\right]^2_{ij} \ge c\|v_x\|^2 \tag{49}$$

for all $v_x \in \mathrm{appl}(z^t)$, where

$$\mathrm{appl}(z^t) = \left\{ v_x \in \mathbb{X} \mid \left[\overline{P}^\top(\nabla g(\overline{x})^* v_x)\overline{P}\right]_{\overline{\gamma}\,\overline{\gamma}} = 0 \right\}. \tag{50}$$

Now, if the S-SOSC does not hold at $\overline{z}$, there exists a nonzero $d \in \mathrm{app}(\overline{z})$ such that

$$\left\langle d, \nabla^2_{xx} L(\overline{z}) d \right\rangle + 2 \sum_{i\in\overline{\alpha}} \sum_{j\in\overline{\gamma}} \frac{-\overline{\lambda}_j}{\overline{\lambda}_i} \left[\overline{P}^\top(\nabla g(\overline{x})^* d)\overline{P}\right]^2_{ij} = 0, \tag{51}$$

where we have used the fact that the SONC holds at $\overline{z}$ to conclude equality (since the left-hand side must be nonnegative by the SONC, and the failure of the S-SOSC implies it is not strictly positive). Since $d \in \mathrm{app}(\overline{z})$, we have $[\overline{P}^\top(\nabla g(\overline{x})^* v_x)\overline{P}]_{\overline{\gamma}\,\overline{\gamma}} = 0$ and $[\overline{P}^\top(\nabla g(\overline{x})^* v_x)\overline{P}]_{\overline{\beta}\overline{\gamma}} = 0$. Then, we know $d \in \mathrm{appl}(z^t)$ as characterized in (50). Put d into (49) and let $t \to 0$, we see that contradicts (51). Thus, the S-SOSC does hold at $\overline{z}$, and we conclude (c)⇒(a). □

In analogy with the discussion for the W-SRCQ, Proposition 5 shows that the W-SOC enjoys a stratified openness property. We next establish a complementary stratified closedness result for the W-SOC.

Proposition 6 *Let $\mathcal{U}$ be an open set in $\mathbb{X} \times \mathbb{Y}$ and let $\widetilde{\mathcal{M}}_{p,q}$ be a stratum of $\mathbb{S}^n$. If the W-SOC holds uniformly at every $z \in \mathcal{U} \cap \widetilde{\mathcal{M}}_{p,q}$ for some p, q, then the W-SOC holds at every $z \in \mathcal{U} \cap \widetilde{\mathcal{M}}_{p',q}$ for all p' such that $0 \le p' \le p$.*

Proof For any p' such that $0 \le p' \le p$ and $z = (x, y) \in \mathcal{U} \cap \widetilde{\mathcal{M}}_{p',q}$ with IED of $G(z)$ given by $(\alpha', \beta', \gamma, p', q, P, \lambda)$, define the matrix

$$\Delta = P \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} P^T,$$

where the block partition corresponds to $[p', p-p', n-p-q, q]$ and I is the identity matrix of order $p-p'$. Then the sequence $z^t = (x, y + t\Delta)$ lies in the stratum $\widetilde{\mathcal{M}}_{p,q}$ for all $t > 0$, and $z^t \to z$ as $t \downarrow 0$. Since $\mathcal{U}$ is open and contains z, there exists $t_0 > 0$ such that $z^t \in \mathcal{U}$ for all $t \in (0, t_0]$. By assumption, the W-SOC holds uniformly at each z^t for $t \in (0, t_0]$. Denote $\alpha = [1, 2, \cdots, p]$ and $\beta = [p+1, p+2, \cdots, n-q]$, we have $\alpha \supseteq \alpha'$, we know $\big(\alpha, \beta, \gamma, p, q, P, \lambda + t\mathrm{diag}(\Delta)\big)$ gives an IED of $G(z^t)$. Moreover, we have $\alpha \supseteq \alpha'$ and $\beta \subseteq \beta'$.

By (44), the uniform validity of the W-SOC at z^t shows there is $c > 0$ such that

$$\big\langle v_x, \nabla^2_{xx} L(z^t) v_x \big\rangle + 2 \sum_{i \in \alpha'} \sum_{j \in \gamma} \frac{-\lambda_j}{\lambda_i} \big[P^\top (\nabla g(x)^* v_x) P\big]^2_{ij} + 2 \sum_{i \in \alpha \setminus \alpha'} \sum_{j \in \gamma} \frac{-\lambda_j}{t} \big[P^\top (\nabla g(x)^* v_x) P\big]^2_{ij} \ge c \|v_x\|^2 \quad (52)$$

for all nonzero $v_x \in \mathrm{appl}(z^t)$, where

$$\mathrm{appl}(z^t) = \big\{ v_x \in \mathbb{X} \mid [P^\top (\nabla g(x)^* v_x) P]_{\beta\beta} = 0,\ [P^\top (\nabla g(x)^* v_x) P]_{\beta\gamma} = 0,\ [P^\top (\nabla g(x)^* v_x) P]_{\gamma\gamma} = 0 \big\}.$$

Then, since $\beta \subseteq \beta'$, it follows that $\mathrm{appl}(z) \subseteq \mathrm{appl}(z^t)$, where

$$\mathrm{appl}(z) = \big\{ v_x \in \mathbb{X} \mid [P^\top (\nabla g(x)^* v_x) P]_{\beta'\beta'} = 0,\ [P^\top (\nabla g(x)^* v_x) P]_{\beta'\gamma} = 0,\ [P^\top (\nabla g(x)^* v_x) P]_{\gamma\gamma} = 0 \big\}.$$

Now, if the W-SOC does not hold at z, there exists some nonzero $d \in \mathrm{appl}(z)$ such that

$$\big\langle d, \nabla^2_{xx} L(z) d \big\rangle + 2 \sum_{i \in \alpha'} \sum_{j \in \gamma} \frac{-\lambda_j}{\lambda_i} \big[P^\top (\nabla g(x)^* d) P\big]^2_{ij} = 0. \quad (53)$$

Since $d \in \mathrm{appl}(z)$, we have $[P^\top (\nabla g(x)^* d) P]_{ij} = 0$ for $i \in \alpha \setminus \alpha'$ and $j \in \gamma$. As $\mathrm{appl}(z) \subseteq \mathrm{appl}(z^t)$, put $v_x = d$ in (52) and let $t \downarrow 0$, we know that contradicts (53). Finally, we conclude that the W-SOC holds at every $z \in \mathcal{U} \cap \widetilde{\mathcal{M}}_{p',q}$ for all p' satisfying $0 \le p' \le p$. □

3.2.3 The regularity across strata

A classical result relating problem-level regularity conditions to solution-level regularity conditions was established in [39]. In Section 3.1, we proved a stratum-restricted counterpart of this equivalence. By further combining it with the equivalences developed in Sections 3.2.1 and 3.2.2, which connect strong-form problem-level conditions at a point to the local uniform validity of their weak-form versions in a neighborhood, we can now establish a comprehensive bridge between the classical regularity conditions and their stratum-restricted counterparts.

Theorem 8 *Let $\overline{x}$ be a local optimal of (1) at which the RCQ holds and $\overline{y} \in M(\overline{x})$. Consider the following statements:*

(a) the S-SOSC and the constraint nondegeneracy hold at $(\overline{x}, \overline{y})$;
(b) every element in $\partial_C F(\overline{x}, \overline{y})$ is nonsingular;
(c) the mapping F is strongly metrically regular at $(\overline{x}, \overline{y})$;
(d) the W-SOC holds uniformly and the W-SRCQ holds at every pair (x, y) sufficiently close to $(\overline{x}, \overline{y})$;
(e) for every (x, y) sufficiently close to $(\overline{x}, \overline{y})$, the manifold differential $dF_{(x,y)}$ is uniformly injective;
(f) for every (x, y) sufficiently close to $(\overline{x}, \overline{y})$, F is stratum-restricted strongly metrically regular on $\widetilde{\mathcal{M}}$ at (x, y), where $\widetilde{\mathcal{M}}$ denotes the stratum containing (x, y).

Then, (a), (b), (c), and (d) are equivalent, and each of them implies (e) and (f).

Proof The equivalence between (a), (b), and (c) follows from [39, Theorem 4.1], while (d)$\Longrightarrow$(e) $\Longleftrightarrow$ (f) is established in Theorem 2. To complete the proof, it remains to show the equivalence between (a) and (d).

(a)$\Rightarrow$(d): This follows directly from Theorem 6 and Theorem 7.

(d)$\Rightarrow$(a): To prove this direction, first note that the validity of the W-SRCQ in a neighborhood of $(\overline{x}, \overline{y})$ implies, by Theorem 6, that the constraint nondegeneracy holds at $(\overline{x}, \overline{y})$. As $\overline{x}$ is feasible and $\overline{y} \in M(\overline{x})$, we know the classical constraint nondegeneracy (Definition 3) holds at $\overline{x}$, thus, the Lagrange multiplier set $\mathcal{M}(\overline{x})$ is a singleton, i.e., $\mathcal{M}(\overline{x}) = \{\overline{y}\}$. By [5, Theorem 3.1 and equation (3.20)] and Remark 9, we conclude that the SONC (as defined in Definition 18) holds at $(\overline{x}, \overline{y})$. Thus, the "Moreover" part of Theorem 7 applies, and we obtain that (d)$\Rightarrow$(a). □

Remark 10 In fact, if statement (e) is reformulated in the following *uniform* version

(e') for every (x, y) sufficiently close to $(\overline{x}, \overline{y})$, the manifold differential $dF_{(x,y)}$ is uniformly injective and positively (or negatively) oriented.

Then it can be made equivalent to (a)–(d) in Theorem 8. Note that we do not provide the precise definitions and proofs here, as they would require introducing additional technical notions that are tangential to our main message. A similar remark applies to statement (f). In any event, the key point of Theorem 8 is to highlight the following perspective: the classical strong-form regularity conditions are equivalent to the local uniform validity of their stratum-restricted counterparts.

Remark 11 If g is affine, then the requirement of uniform validity on a neighborhood can be weakened to pointwise validity on that neighborhood. Indeed, in this case, by noting that (52) holds, so the relevant results follows immediately.

Remark 12 From a primal–dual perspective, one may wonder why statement (d) requires a uniform version of the W-SOC, whereas no uniformity is imposed for the W-SRCQ. We provide the following intuitive explanation; a rigorous analysis is deferred to future work. The W-SRCQ is driven by the affine structure of the linearized constraint mapping $\nabla g(x)^*$; as noted in Remark 11, this structure effectively yields uniformity once the condition holds locally. By contrast, the W-SOC, which can be viewed as a dual analogue of the W-SRCQ, does not retain this affine character, due to the spectral nonlinearity induced by the non-polyhedral cone $\mathbb{S}^n_+$, and therefore a uniform formulation becomes essential.

4 A globalized Gauss–Newton method based on stratification

In this section, we leverage the stratification framework developed above, which provides a principled way to handle the nonsmoothness of KKT mapping F given in (2), to design a Newton-type method that is globally convergent and enjoys fast local convergence for solving the KKT system of (1). Specifically, to compute a KKT pair of (1), we consider the global least-squares reformulation

$$\begin{aligned} \min \quad & \varphi(z) := \frac{1}{2}\|F(z)\|^2 \\ \text{s.t.} \quad & z \in \mathbb{X} \times \mathbb{S}^n. \end{aligned} \tag{54}$$

Because the nonsmoothness of F is resolved only after restricting to a fixed stratum, any global method must be able to move between strata of different dimensions. Accordingly, our algorithm combines three complementary mechanisms:

- *Move along a stratum:* take a Levenberg–Marquardt step on the current stratum, implemented via a line-search-compatible retraction to ensure descent of φ.
- *Escape to higher-dimensional strata:* exploit normal directions to the current stratum, where an explicit step can be computed and accepted without a line search.
- *Transition to lower-dimensional strata:* when a sufficiently small eigenvalue is detected, apply a correction step to identify a lower-dimensional stratum with a smaller KKT residual.

Under mild assumptions, we shall establish global convergence of the proposed method and prove fast local convergence. In particular, the regularity requirement guaranteeing local quadratic convergence is strictly weaker than the regularity assumptions typically imposed in existing Newton-type approaches that achieve superlinear or quadratic rates.

4.1 Algorithm

In this subsection, we outline the proposed algorithm at a conceptual level, postponing precise definitions to the subsequent discussion. We refer to the method as the *Stratified Gauss–Newton method with correction* (SGN; see Algorithm 1). At each iteration, the current iterate z^k may first be corrected to $\widehat{z^k}$, with the goal of moving it onto a stratum where the desired regularity properties are more likely to hold. Starting from $\widehat{z^k}$, we then compute a stratum LM-normal step, denoted by $\texttt{SLMN}(\widehat{z^k})$ (see Algorithm 2), and apply an acceptance test based on the decrease of the merit function φ. This acceptance mechanism ensures that the sequence $\{\varphi(z^k)\}$ is monotonically decreasing, thereby preventing potential increase in φ due to the correction step.

Algorithm 1 Stratified Gauss–Newton Method with correction (SGN)

1: Initialize $z^0 \in \mathbb{X} \times \mathbb{S}^n$, correction threshold $\delta > 0$ and tolerance $\epsilon > 0$.
2: **for** $k = 0, 1, \ldots$ until z^k satisfies $s(z^k, \varphi) \le \epsilon$ (Definition 21) **do**
3: Compute $\widehat{z^k}$ from z^k by (67) with δ.
4: **if** $\varphi\left(\texttt{SLMN}(\widehat{z^k})\right) \le \varphi(z^k)$ **then**
5: $z^{k+1} \leftarrow \texttt{SLMN}(\widehat{z^k})$
6: **else**
7: $z^{k+1} \leftarrow \texttt{SLMN}(z^k)$
8: **end if**
9: **end for**

Algorithm 1 is built upon three main components: a stationarity-measure step $s(\cdot, \varphi)$, a stratum LM-normal step $\texttt{SLMN}(\cdot)$, and a correction step $\widehat{\cdot}$. These subroutines will be introduced and analyzed in detail in the following subsections.

4.1.1 Directional stationarity and stationarity measure

A suitable stopping rule is essential for the design of an iterative algorithm in practice. Conventionally, the norm of the gradient of the merit function is used as a stopping criterion. However, due to the nonsmoothness of φ, a more careful analysis is required, leading us to the following notion of *directional stationarity*.

We first recall the standard definition of directional stationarity in nonsmooth optimization (see, e.g., [18, Section 8.2]).

Definition 20 Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and directionally differentiable. A point $z^* \in \mathbb{R}^n$ is called a *directionally stationary point* of ϕ if

$$\phi'(z^*; v) \ge 0 \quad \forall\, v \in \mathbb{R}^n.$$

For brevity, we also refer to such a point as *D-stationary*.

We therefore begin by computing the directional derivative of the merit function, which is precisely the objective function of (54),

$$\varphi(z) = \frac{1}{2}\|F(z)\|^2. \tag{55}$$

In the previous sections we wrote a tangent vector $v \in \mathcal{T}_z\widetilde{\mathcal{M}}_{p,q} \subseteq \mathbb{X} \times \mathbb{S}^n$ as $v = (v_x, v_y)$. In the present discussion we shall consider an arbitrary direction $v \in \mathbb{X} \times \mathbb{S}^n$ (not necessarily tangent to a fixed stratum), and we continue to write $v = (v_x, v_y)$ for notational convenience.

Fix $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$ and let $G(z)$ admit an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$. Recall the orthogonal decomposition

$$\mathbb{S}^n = \mathcal{T}_{G(z)}\mathcal{M}_{p,q} \;\oplus\; \mathcal{N}_{G(z)}\mathcal{M}_{p,q}, \tag{56}$$

where

$$\mathcal{T}_{G(z)}\mathcal{M}_{p,q} = \left\{\, P\widetilde{H}P^\top \mid \widetilde{H} \in \mathbb{S}^n,\ \widetilde{H}_{\beta\beta} = 0 \,\right\}$$

and

$$\mathcal{N}_{G(z)}\mathcal{M}_{p,q} = \left\{H \in \mathbb{S}^n \mid \text{ only } (P^\top H P)_{\beta\beta} \text{ can be nonzero } \right\}.$$

The next lemma lifts this decomposition to $\mathbb{X} \times \mathbb{S}^n$ in a form compatible with the coordinate map ϕ_z introduced in (23).

Lemma 4 *For an arbitrary $z \in \mathbb{X} \times \mathbb{S}^n$ with $G(z)$ having an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$ and $v = (v_x, v_y) \in \mathbb{X} \times \mathbb{S}^n$, under the isomorphism ϕ_z defined in* (23)*, we have a decomposition $\phi_z(v) = u_1 + u_2$ such that*

$$u_1 = \begin{bmatrix} v_x \\ H_1 \end{bmatrix} \in \mathbb{X} \times \mathcal{T}_{G(z)}\mathcal{M}_{p,q},\ \text{where } H_1 \in \mathcal{T}_{G(z)}\mathcal{M}_{p,q} = \left\{H = P\widetilde{H}P^\top \mid \widetilde{H} \in \mathbb{S}^n \text{ and } \widetilde{H}_{\beta\beta} = 0\right\},$$

$$u_2 = \begin{bmatrix} 0 \\ H_2 \end{bmatrix} \in \{0\} \times \mathcal{N}_{G(z)}\mathcal{M}_{p,q},\ \text{where } H_2 \in \mathcal{N}_{G(z)}\mathcal{M}_{p,q} = \left\{H = P\widetilde{H}P^\top \mid \widetilde{H} \in \mathbb{S}^n \text{ and only } \widetilde{H}_{\beta\beta} \text{ can be nonzero}\right\}.$$

Moreover, defining $\mathcal{N}_z\widetilde{\mathcal{M}}_{p,q} := \phi_z^{-1}(\{0\} \times \mathcal{N}_{G(z)}\mathcal{M}_{p,q})$, we can decompose $\mathbb{X} \times \mathbb{S}^n$ as

$$\mathbb{X} \times \mathbb{S}^n = \mathcal{T}_z\widetilde{\mathcal{M}}_{p,q} \oplus \mathcal{N}_z\widetilde{\mathcal{M}}_{p,q}$$

and get $v = v_1 + v_2$ with $v_1 \in \mathcal{T}_z\widetilde{\mathcal{M}}_{p,q}, v_2 \in \mathcal{N}_z\widetilde{\mathcal{M}}_{p,q}$.

Proof It is easy to check that the isomorphism ϕ_z in Lemma 1 can actually be extended to be an isomorphism from $\mathbb{X}\times\mathbb{S}^n$ to $\mathbb{X}\times\mathbb{S}^n$ using the same expression, i.e.,

$$\begin{aligned}\phi_z : \mathbb{X}\times\mathbb{S}^n &\to \mathbb{X}\times\mathbb{S}^n\\ (v_x, v_y) &\mapsto (v_x, H = \nabla g(x)^*(v_x) + v_y)\end{aligned}$$

and

$$\begin{aligned}\phi_z^{-1} : \mathbb{X}\times\mathbb{S}^n &\to \mathbb{X}\times\mathbb{S}^n\\ (v_x, H) &\mapsto (v_x, v_y = H - \nabla g(x)^*(v_x))\end{aligned}$$

are a pair of isomorphism from $\mathbb{X}\times\mathbb{S}^n$ to $\mathbb{X}\times\mathbb{S}^n$. With the decomposition (56), we can decompose $\mathbb{X}\times\mathbb{S}^n$ as

$$\mathbb{X}\times\mathbb{S}^n = (\mathbb{X}\times\mathcal{T}_{G(z)}\mathcal{M}_{p,q}) \oplus (\{0\}\times\mathcal{N}_{G(z)}\mathcal{M}_{p,q}). \tag{57}$$

Therefore, $\phi_z(v)\in\mathbb{X}\times\mathbb{S}^n$ can be decomposed as $\phi_z(v) = u_1 + u_2$ such that $u_1 = \begin{bmatrix} v_x \\ H_1\end{bmatrix} \in \mathbb{X}\times\mathcal{T}_{G(z)}\mathcal{M}_{p,q}$, and $u_2 = \begin{bmatrix} 0 \\ H_2\end{bmatrix} \in \{0\}\times\mathcal{N}_{G(z)}\mathcal{M}_{p,q}$.

Applying ϕ_z^{-1} to both sides of $\phi_z(v)\in\mathbb{X}\times\mathbb{S}^n$, we get $v = v_1 + v_2$ with $v_1 = \phi_z^{-1}(u_1), v_2 = \phi_z^{-1}(u_2)$. By the isomorphism in Lemma 1 and the definition of $\mathcal{N}_z\widetilde{\mathcal{M}}_{p,q}$, we know $v_1\in\mathcal{T}_z\widetilde{\mathcal{M}}_{p,q}$ and $v_2\in\mathcal{N}_z\widetilde{\mathcal{M}}_{p,q}$. □

Remark 13 For readers familiar with Riemannian geometry, the above construction can be interpreted as twisting the Riemannian metric on $\mathbb{X}\times\mathbb{S}^n$ via the map g. Consequently, $\mathcal{N}_z\widetilde{\mathcal{M}}$ is indeed normal to $\mathcal{T}_z\widetilde{\mathcal{M}}$ with respect to this twisted metric. This perspective also highlights a general guiding principle: an appropriate choice of Riemannian metric can induce a direct-sum decomposition that aligns naturally with the underlying structure of the problem.

Now we can compute the directional derivative $\varphi'(z;v)$ for $z=(x,y)\in\mathbb{X}\times\mathbb{S}^n$ and an arbitrary direction $v\in\mathbb{X}\times\mathbb{S}^n$. The next proposition provides an explicit formula. Its proof is a straightforward calculation and is therefore deferred to Appendix C.

Proposition 7 *For an arbitrary $z\in\mathbb{X}\times\mathbb{S}^n$ with $G(z)$ having an IED $(\alpha,\beta,\gamma,p,q,P,\lambda)$ and $v=(v_x,v_y)\in\mathbb{X}\times\mathbb{S}^n$, the directional derivative of φ at z along v can be expressed as*

$$\varphi'(z,v) = \left\langle dF_z^*(F(z)), \begin{bmatrix} v_x \\ H_1\end{bmatrix}\right\rangle + \left\langle \nabla g(x)^*(F_1(z)), \Pi_{\mathbb{S}^n_-}(H_2)\right\rangle + \left\langle \nabla g(x)^*(F_1(z)) + F_2(z), \Pi_{\mathbb{S}^n_+}(H_2)\right\rangle, \tag{58}$$

where $F_1(z) = \nabla_x L(x,y) = \nabla f(x) + \langle y, \nabla g(x)\rangle$, $F_2(z) = -g(x) + \Pi_{\mathbb{S}^n_+}(G(z)) = -g(x) + \Pi_{\mathbb{S}^n_+}(g(x)+y)$ and H_1 and H_2 are defined in Lemma 4.

We can use the calculation in Proposition 7 to give a characterization of D-stationary point of φ. First, we need to define two directions. Let $z=(x,y)\in\widetilde{\mathcal{M}}_{p,q}$ with an IED $(\alpha,\beta,\gamma,p,q,P,\lambda)$ for $G(z)$. Recall the decomposition of $\mathbb{S}^n$ as

$$\mathbb{S}^n = \mathcal{T}_{G(z)}\mathcal{M}_{p,q}\oplus\mathcal{N}_{G(z)}\mathcal{M}_{p,q},$$

where the tangent and normal spaces are given by

$$\mathcal{T}_{G(z)}\mathcal{M}_{p,q} = \left\{H = P\widetilde{H}P^\top \mid \widetilde{H}\in\mathbb{S}^n,\ \widetilde{H}_{\beta\beta} = 0\right\}$$

and

$$\mathcal{N}_{G(z)}\mathcal{M}_{p,q} = \left\{H\in\mathbb{S}^n \mid \text{ only } (P^\top HP)_{\beta\beta} \text{ can be nonzero }\right\}.$$

We define the projections of a matrix $H\in\mathbb{S}^n$ onto these spaces as follows:

$$\pi^1_{z,\mathbb{S}^n}(H) = P\begin{bmatrix}\widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \widetilde{H}_{\alpha\gamma}\\ \widetilde{H}_{\beta\alpha} & 0 & \widetilde{H}_{\beta\gamma}\\ \widetilde{H}_{\gamma\alpha} & \widetilde{H}_{\gamma\beta} & \widetilde{H}_{\gamma\gamma}\end{bmatrix}P^\top \text{ and } \pi^2_{z,\mathbb{S}^n}(H) = P\begin{bmatrix}0 & 0 & 0\\ 0 & \widetilde{H}_{\beta\beta} & 0\\ 0 & 0 & 0\end{bmatrix}P^\top, \tag{59}$$

where $\widetilde{H} = \begin{bmatrix}\widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \widetilde{H}_{\alpha\gamma}\\ \widetilde{H}_{\beta\alpha} & \widetilde{H}_{\beta\beta} & \widetilde{H}_{\beta\gamma}\\ \widetilde{H}_{\gamma\alpha} & \widetilde{H}_{\gamma\beta} & \widetilde{H}_{\gamma\gamma}\end{bmatrix} = P^\top HP$. Now, we define the normal directions $W_1(z)$ and $W_2(z)\in\mathcal{N}_{G(z)}\mathcal{M}_{p,q}$ as follows:

$$W_1(z) = \Pi_{\mathbb{S}^n_-}\left(-\pi^2_{z,\mathbb{S}^n}\left(\nabla g(x)^*(F_1(z))\right)\right), \tag{60}$$

and

$$W_2(z) = \Pi_{\mathbb{S}^n_+}\left(-\pi^2_{z,\mathbb{S}^n}\left(\nabla g(x)^*(F_1(z)+F_2(z))\right)\right). \tag{61}$$

Proposition 8 *Let $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$, then z is a D-stationary point of φ if and only if all the following three conditions hold:*

1. $dF_z^*(F(z)) = 0$;
2. $W_1(z) := \Pi_{\mathbb{S}^n_-}(-\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*(F_1(z)))) = 0$, *i.e.* $\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*(F_1(z))) \in \mathbb{S}^n_-$;
3. $W_2(z) := \Pi_{\mathbb{S}^n_+}\big(-\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*(F_1(z)) + F_2(z))\big) = 0$, *i.e.* $\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*(F_1(z)) + F_2(z)) \in \mathbb{S}^n_+$.

Proof Noting that $dF_z^*(F(z)) \in \mathcal{T}_z\widetilde{\mathcal{M}}$, we have $\left\langle dF_z^*(F(z)), \begin{bmatrix} v_x \\ H_1 \end{bmatrix} \right\rangle \geq 0$ for all $\begin{bmatrix} v_x \\ H_1 \end{bmatrix} \in \mathcal{T}_z\widetilde{\mathcal{M}}$ if and only if $dF_z^*(F(z)) = 0$.

Checking the definition of $\mathbb{S}^n_+$, it's easy to know that $\left\langle \nabla g(x)^*(F_1(z)), \Pi_{\mathbb{S}^n_-}(H_2) \right\rangle \geq 0$ if and only if the image of $\nabla g(x)^*(F_1(z))$ projected onto $\mathcal{N}_{G(z)}\mathcal{M}$ is in $\mathbb{S}^n_-$, i.e. $\Pi_{\mathbb{S}^n_+}\big(\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^* F_1(z))\big) = 0$, i.e. $W_1(z) = 0$. A similar argument works for $W_2(z)$. □

Remark 14 One should observe that Statement 1 in the above proposition is exactly the characterization of a stationary point of the stratum-restricted version of (54).

In view of Proposition 8, a straightforward idea is to measure the D-stationarity of a point by these three directions.

Definition 21 For $z \in \mathbb{X} \times \mathbb{S}^n$, we define the *D-stationarity measurement* of φ at z as the function

$$\begin{aligned} s(\cdot) : \mathbb{X} \times \mathbb{S}^n &\to \mathbb{R} \\ z &\mapsto \max\{\|W_1(z)\|, \|W_2(z)\|, \|v^{\mathrm{LM}}(z)\|\}, \end{aligned}$$

where $W_1(z)$ and $W_2(z)$ are defined in (60) and (61), and $v^{\mathrm{LM}}(z)$ is the stratum-LM direction defined in (65).

Remark 15 This D-stationarity measurement function is not globally continuous. Its value might mutate when some eigenvalue of $G(z)$ goes to 0 since the definition of W_1, W_2 and v all depend on the tangent space $\mathcal{T}_z\widetilde{\mathcal{M}}_{p,q}$, whose dimension will mutate in this case.

4.1.2 `SLMN` *descent step*

To propose a global algorithm for the unconstrained problem (54), finding a descent direction is at the core. Motivated by the stratified structure of $\mathbb{X} \times \mathbb{S}^n$, we shall exploit two types of directions: tangential directions that move within the current stratum and reduce the residual, and normal directions that facilitate transitions across strata. This leads to the *stratum LM-normal step*, denoted by `SLMN`. At an iterate z^k, the `SLMN` step combines a stratum-tangential component with suitably chosen normal components so as to ensure sufficient decrease of the merit function $\varphi(z) = \frac{1}{2}\|F(z)\|^2$ while accommodating the nonsmoothness of F. We outline the construction below and then present the formal procedure; see Algorithm 2 for details.

Tangential direction. In line with our stratified approach, we compute at each iterate z^k a step that reduces the residual while remaining tangent to the current stratum. Specifically, we consider the stratum-restricted subproblem

$$\begin{aligned} \min \quad & \varphi(z) \\ \text{s.t.} \quad & z \in \widetilde{\mathcal{M}}_{p,q}, \end{aligned} \tag{62}$$

where the stratum $\widetilde{\mathcal{M}}_{p,q}$ is determined by the current iterate. A Gauss–Newton step is obtained by minimizing the quadratic model of φ at z^k over the tangent space:

$$\begin{aligned} \min \quad & \varphi_k(v) := \frac{1}{2}\big\|F(z^k) + dF(z^k)[v]\big\|^2 \\ \text{s.t.} \quad & v \in \mathcal{T}_{z^k}\widetilde{\mathcal{M}}_{p,q}. \end{aligned} \tag{63}$$

This convex quadratic program is equivalent to the normal equation

$$\big(dF(z^k)\big)^* dF(z^k)[v] = -\big(dF(z^k)\big)^* F(z^k), \qquad v \in \mathcal{T}_{z^k}\widetilde{\mathcal{M}}_{p,q}, \tag{64}$$

where $(dF(z^k))^*$ denotes the adjoint operator. By Theorem 2, if the W-SOC and the W-SRCQ hold at z^k, then (64) admits a unique solution. Since our goal is global convergence, we further incorporate the Levenberg–Marquardt regularization to prevent degeneracy of $(dF(z^k))^* dF(z^k)$ when far from the solution.

Definition 22 The *stratum-LM direction* of φ or F at $z \in \mathbb{X} \times \mathbb{S}^n$ is defined to be

$$v^{\mathrm{LM}}(z) = -(\mu(z)I + dF_z^* dF_z)^{-1} dF_z^* F(z), \tag{65}$$

where $\mu(z) = \|F(z)\|^2$ is the regularizing term.

Remark 16 Building upon the classical Levenberg–Marquardt methods (see, e.g., [45]), it is straightforward to construct a local algorithm that solves (62) quadratically, provided that the algorithm starts sufficiently close to a KKT pair $\overline{z}$, and the W-SOC and W-SRCQ hold at $\overline{z}$. This result will be incorporated into the local analysis of our global algorithm, as detailed in Proposition 12.

Once the stratum-LM direction is computed, the corresponding step is obtained by $z^{k+1} = R_{z^k}(v^k)$, where R_{z^k} is a retraction map that projects $z^k + tv^k$ back onto the current stratum $\widetilde{\mathcal{M}}_{p,q}$. We will discuss this retraction operation in more detail below.

For a general manifold $\mathcal{M}$, a classical candidate for retraction is the Riemannian exponential map, which was used in [38] to establish the local quadratic convergence of the Riemannian Newton method. However, computing the Riemannian exponential directly is often intractable due to the need to solve geodesic equations explicitly. A more practical alternative is the metric projection onto $\mathcal{M}$, which is well-defined and locally smooth, provided the projection is unique [42,1].

In this work, rather than directly applying the metric projection on $\widetilde{\mathcal{M}}_{p,q}$, we introduce an induced retraction. Specifically, we first define a retraction on $\mathcal{M}_{p,q}$ using the metric projection, and then extend it to $\widetilde{\mathcal{M}}_{p,q}$. This approach leverages the computational efficiency of the metric projection on $\mathcal{M}_{p,q}$, which can be computed via truncated SVD, instead of projecting onto the more complicated manifolds $\widetilde{\mathcal{M}}_{p,q}$. It should be noted that while a direct projection-based retraction is feasible, the induced retraction also offers computational benefits, particularly for evaluating $\varphi(\cdot)$, which is essential for line search procedures.

We begin by introducing the retraction for $\mathcal{M}_{p,q}$ and then extend it to $\widetilde{\mathcal{M}}_{p,q}$. The retraction $R_{\mathcal{M}} : \mathcal{T}\mathcal{M} \to \mathcal{M}$ on $\mathcal{M} = \mathcal{M}_{p,q}$ is given by:

$$\Pi_{\mathcal{M}}(A + H) := \mathrm{argmin}_{M \in \mathcal{M}} \|M - (A + H)\|_F^2,$$

where $\Pi_{\mathcal{M}}$ is the metric projection onto $\mathcal{M}_{p,q}$. It is well-known that $\Pi_{\mathcal{M}}$ can be computed explicitly. If $A \in \mathcal{M}_{p,q}$ has an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$ and $A+H$ has an IED $(\alpha', \beta', \gamma', p', q', P', \lambda')$, with $p' \geq p$ and $q' \geq q$ for sufficiently small $\|H\|$, the projection is given by:

$$\Pi_{\mathcal{M}}(A + H) = P' \Lambda' (P')^\top,$$

where $\Lambda' = \begin{bmatrix} \mathrm{diag}(\lambda'_\alpha) & 0 & 0 \\ 0 & 0_{\beta\beta} & 0 \\ 0 & 0 & \mathrm{diag}(\lambda'_\gamma) \end{bmatrix}$.

By [1, Lemma 3.1], $R_{\mathcal{M}}$ is a smooth retraction since $\mathcal{M}_{p,q}$ is a C^∞ smooth submanifold of $\mathbb{S}^n$. Next, we define the retraction $R_{\widetilde{\mathcal{M}}} : \mathcal{T}\widetilde{\mathcal{M}} \to \widetilde{\mathcal{M}}$ for $\widetilde{\mathcal{M}} = \widetilde{\mathcal{M}}_{p,q}$ as follows: for a tangent vector $v = (v_x, v_y) \in \mathcal{T}_z\widetilde{\mathcal{M}}$ at $z = (x, y) \in \widetilde{\mathcal{M}}$,

$$R_{\widetilde{\mathcal{M}}}(z, v) := (x + v_x, R_{\mathcal{M}}(A, H) - g(x + v_x)), \tag{66}$$

where $A = G(z) = g(x) + y \in \mathcal{M}$ since $z = (x, y) \in \widetilde{\mathcal{M}}$, and $H = \nabla G(z)^*(v) = \nabla g(x)^*(v_x) + v_y \in \mathcal{T}_A\mathcal{M}$. We also define the notation $R_{\widetilde{\mathcal{M}},z}(v) := R_{\widetilde{\mathcal{M}}}(z, v)$, and simplify to $R_z(v)$ when $\widetilde{\mathcal{M}}$ is clear from the context.

Remark 17 For readers familiar with fiber bundles, the retraction $R_{\widetilde{\mathcal{M}}}$ on $\widetilde{\mathcal{M}}$ can be interpreted as performing retraction along both the fiber and base directions with respect to the fiber bundle induced by the submersion $G : \widetilde{\mathcal{M}} \to \mathcal{M}$.

Proposition 9 *$R_{\widetilde{\mathcal{M}}}$ is a smooth retraction.*

Proof Verifying the definition is routine, and the smoothness follows from the smoothness of $R_{\mathcal{M}}$. □

Normal direction: The stratum LM direction remains within the current stratum, but to move between strata, we shall follow normal directions to escape from a "bad" stratum that does not contain a solution of (54). These directions come from (60) and (61) we defined above. Since $W_i(z) \in \mathcal{N}_{G(z)}\mathcal{M}_{p,q}$, then by the definition of ϕ_z in (23) and definition of $\mathcal{N}_z\widetilde{\mathcal{M}}_{p,q}$ in Lemma 4, an explicit calculation shows

$$\begin{bmatrix} 0 \\ W_i(z) \end{bmatrix} = \phi_z^{-1}\left(\begin{bmatrix} 0 \\ W_i(z) \end{bmatrix}\right) \in \mathcal{N}_z\widetilde{\mathcal{M}}_{p,q} \subseteq \mathbb{X} \times \mathbb{S}^n,$$

and we can discuss the step $z + \begin{bmatrix} 0 \\ W_i(z) \end{bmatrix}$. These normal directions $\begin{bmatrix} 0 \\ W_i(z) \end{bmatrix}$ are used to transition to higher-dimensional strata. The following lemma characterizes how $\varphi(z)$ varies along these directions.

Lemma 5 *Given some $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$. If $W_1(z) \neq 0$, then it implies that $\|\nabla g(x) W_1(z)\| \neq 0$ and we have*

$$\operatorname*{argmin}_{t \geq 0} \varphi\left(z + t \begin{bmatrix} 0 \\ W_1(z) \end{bmatrix}\right) = \frac{\|W_1(z)\|^2}{\|\nabla g(x) W_1(z)\|^2}$$

and

$$\varphi(z) - \min_{t \geq 0} \varphi\left(z + t \begin{bmatrix} 0 \\ W_1(z) \end{bmatrix}\right) = \frac{1}{2} \frac{\|W_1(z)\|^4}{\|\nabla g(x) W_1(z)\|^2}.$$

Similarly, if $W_2(z) \neq 0$, we have

$$\operatorname*{argmin}_{t \geq 0} \varphi\left(z + t \begin{bmatrix} 0 \\ W_2(z) \end{bmatrix}\right) = \frac{\|W_2(z)\|^2}{\|W_2(z)\|^2 + \|\nabla g(x) W_2(z)\|^2}$$

and

$$\varphi(z) - \min_{t \geq 0} \varphi\left(z + t \begin{bmatrix} 0 \\ W_2(z) \end{bmatrix}\right) = \frac{1}{2} \frac{\|W_2(z)\|^4}{\|W_2(z)\|^2 + \|\nabla g(x) W_2(z)\|^2}.$$

Moreover, suppose Ω is a sufficiently large compact domain e.g. containing the whole area the algorithm is running, denoting the Lipschitz constant of g on Ω by $L_{g,\Omega}$ (due to the smoothness of g on $\mathbb{X}$), we have

$$\varphi(z) - \min_{t \geq 0} \varphi\left(z + t \begin{bmatrix} 0 \\ W_1(z) \end{bmatrix}\right) \geq \frac{1}{2L_{g,\Omega}^2} \|W_1(z)\|^2 \text{ or } \varphi(z) - \min_{t \geq 0} \varphi\left(z + t \begin{bmatrix} 0 \\ W_2(z) \end{bmatrix}\right) \geq \frac{1}{2(L_{g,\Omega}^2 + 1)} \|W_2(z)\|^2,$$

if $W_1(z) \neq 0$ or $W_2(z) \neq 0$, respectively.

Since $\varphi\left(z + t \begin{bmatrix} 0 \\ W_i(z) \end{bmatrix}\right)$ is a quadratic function of t, the conclusion is obtained by direct computation again. We will also put its proof in Appendix C.

With the construction of the directions, the following is the precise form of our descent step algorithm $\texttt{SLMN}(z^k)$.

Algorithm 2 Stratum LM-normal step (SLMN)

1: Given $z^k \in \mathcal{X} \times \mathcal{Y}$, $\eta \in (\frac{1}{2}, 1), \rho \in (0, 1)$.

2: **Switch:**

- if $W_1^k := W_1(z^k) \neq 0$ (see (60)), compute:

$$z_1^{k+1} = z^k + \frac{\|W_1^k\|^2}{\|\nabla g(x^k) W_1^k\|^2} \begin{bmatrix} 0 \\ W_1^k \end{bmatrix};$$

- if $W_2^k := W_2(z^k) \neq 0$ (see (61)), compute:

$$z_2^{k+1} = z^k + \frac{\|W_2^k\|^2}{\|W_2^k\|^2 + \|\nabla g(x^k) W_2^k\|^2} \begin{bmatrix} 0 \\ W_2^k \end{bmatrix};$$

- perform manifold Armijo line search along $v^k := v^{\mathrm{LM}}(z^k)$ (see (65)), i.e. find the smallest $j_k \in \mathbb{Z}_+$ such that

$$\varphi(R_{z^k}(\rho^{j_k} v^k)) - \varphi(z^k) \leq \eta \cdot \rho^{j_k} \varphi'(z^k, v^k)$$

and set $z_3^{k+1} = R_{z^k}(\rho^{j_k} v^k)$.

3: Return

$$z^{k+1} = \texttt{SLMN}(z^k) \leftarrow \arg\min \left\{ \varphi(z) \mid z \in \{z_1^{k+1}, z_2^{k+1}, z_3^{k+1}\} \right\},$$

where non-existent points are excluded from the comparison.

In Algorithm 2, z_1^{k+1} and z_2^{k+1} are computed based on Lemma 5, which results in a decrease in the function value. The line search step clearly leads to a decrease in φ, provided that j_k exists, which is proved in Proposition C.1. Therefore, we can conclude that SLMN always results in a decrease of φ at non-D-stationary points.

4.1.3 Correction strategy

Algorithm 2 provides a clear direction for navigating within an arbitrary stratum of $\mathbb{X} \times \mathbb{S}^n$. However, if the algorithm iterates to a point near the boundary of a stratum, the differential information about the adjacent stratum may not be readily available. Fortunately, by leveraging a correction strategy for stratum identification proposed in [19] and inspired by the globalization approach for smooth functions on stratified spaces developed in [32], we are able to steer the iterates toward the stratum containing the solution, provided the solution satisfies a suitable regularity condition.

To be specific, for the correction strategy, let $\delta > 0$ be fixed and let $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$ admit an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$ of $G(z)$. Following [19], we introduce the following corrected point:

$$\widehat{z} := \left(x,\ y - \sum_{i \in \theta(z)} \lambda_i(G(z))\, P_i P_i^\top \right), \tag{67}$$

where the index set is defined by

$$\theta(z) = \big\{ i \mid |\lambda_i(G(z))| \le \delta \big\}.$$

It should be noted that one can in fact view the operator $\widehat{\cdot}$ as a locally smooth mapping from $\mathbb{X} \times \mathbb{S}^n$ to itself, which follows directly from the smoothness properties of the Löwner operator.

The following lemma guarantees that the decrement controlled in the above two parts are the main part so that our arguments can work.

Lemma 6 *Suppose a sequence $\{z^k\}$ convergences to $z^\infty \in \widetilde{\mathcal{M}}$. Then when k is large enough, we have*

$$\|\widehat{z^k} - z^k\| \le O(\|z^k - z^\infty\|)$$

and hence

$$\|\widehat{z^k} - z^\infty\| \le O(\|z^k - z^\infty\|).$$

Proof Denote $\alpha^k = \alpha(G(z^k))$, $\widehat{\alpha^k} = \alpha(G(\widehat{z^k}))$ and $\alpha^\infty = \alpha(G(z^\infty))$. Similar for β and γ. Then we prove this lemma by direct calculation. First, note that $\widehat{z^k}$ and z^k have the same x-part. So we have

$$\|\widehat{z^k} - z^k\| = \|\big(x^k, G(\widehat{z^k}) - g(x^k)\big) - \big(x^k, G(z^k) - g(x^k)\big)\| = \|G(\widehat{z^k}) - G(z^k)\| \tag{68}$$

Then we assume that $G(z^k)$ has an eigen-decomposition

$$G(z^k) = P^k \begin{bmatrix} \Lambda^k_{\alpha^k} & & \\ & 0_{\beta^k} & \\ & & \Lambda^k_{\gamma^k} \end{bmatrix} (P^k)^\top \quad \text{and} \quad G(z^\infty) = P^\infty \begin{bmatrix} \Lambda^\infty_{\alpha^\infty} & & \\ & 0_{\beta^\infty} & \\ & & \Lambda^\infty_{\gamma^\infty} \end{bmatrix} (P^\infty)^\top.$$

By definition, $G(\widehat{z^k})$ is $G(z^k)$ with some of entries of $\Lambda^k_{\alpha^k}$ and $\Lambda^k_{\gamma^k}$ changed to be 0. So their difference will be

$$G(z^k) - G(\widehat{z^k}) = P^k \begin{bmatrix} \widetilde{\Lambda^k}_{\alpha^k} & & \\ & 0_{\beta^k} & \\ & & \widetilde{\Lambda^k}_{\gamma^k} \end{bmatrix} (P^k)^\top,$$

with some of entries of $\widetilde{\Lambda^k}_{\alpha^k}$ and $\widetilde{\Lambda^k}_{\gamma^k}$ are nonzero. Since $G(z^k) \to G(z^\infty)$, we have $\alpha^k \ge \widehat{\alpha^k} = \alpha^\infty$, $\gamma^k \ge \widehat{\gamma^k} = \gamma^\infty$. And when k is large enough, the eigenvalues belonging to $\widehat{\alpha^k}$-part will be larger than δ since they are converging to positive eigenvalues larger than $\delta(z^\infty) > \delta$ and those in $\alpha^k \setminus \widehat{\alpha^k}$ will be corrected to be 0 since they are converging to 0. Similarly, the eigenvalues of $G(z^k)$ corrected to be 0 are exactly those are converging to 0. Hence $\widetilde{\Lambda^k}$ will concentrate in $\beta^\infty (= \widehat{\beta^k})$-part and we have

$$\begin{aligned} \left\| G(z^k) - G(\widehat{z^k}) \right\| = \left\| \widetilde{\Lambda^k} \right\| &= \left\| \Lambda_{\widehat{\beta^k}}(G(z^k)) - \Lambda_{\beta^\infty}(G(z^\infty)) \right\| \\ &\le \left\| \Lambda(G(z^k)) - \Lambda(G(z^\infty)) \right\| \le \left\| G(z^k) - G(z^\infty) \right\| \le O(\|z^k - z^\infty\|). \end{aligned} \tag{69}$$

Putting (68) and (69) together, we get the first claim. The second follows by

$$\|\widehat{z^k} - z^\infty\| \le \|z^k - \widehat{z^k}\| + \|z^k - z^\infty\|.$$

This completes the proof. □

4.2 Convergence analysis

In this subsection, we present the convergence analysis of the proposed Algorithm 1. The following result establishes its global convergence, whose proof is provided in Appendix C. To establish this, we introduce the following definition of the lower modulus of nonzero eigenvalues: for every $z \in \mathbb{X} \times \mathbb{S}^n$,

$$\delta(z) := \min\{|\lambda(G(z))| : \lambda(G(z)) \neq 0\}. \tag{70}$$

Theorem 9 *Taking tolerance to be $\epsilon = 0$ and starting from an arbitrary initial point z^0, Algorithm 1 will stop in finite steps or we get an infinite sequence. For the finite steps case, Algorithm 1 will stop at a D-stationary point of $\varphi(z) = \frac{1}{2}\|F(z)\|^2$. For the infinite sequence case, any accumulation point z^∞ satisfying $\delta(z^\infty) > \delta$ of the sequence is a D-stationary point of $\varphi(z)$.*

Remark 18 An issue in Theorem 9 is that the hyper-parameter δ must satisfy a condition that involves the unknown accumulation point. This restrictive requirement can be bypassed by introducing a *partial correction* strategy, which considers all possible correction points rather than only $\widehat{z^k}$. A similar approach is utilized by Olikier, Gallivan, and Absil in [32] for optimizing smooth functions over stratified spaces. Moreover, this strategy lends itself to a parallel structure, ensuring computational efficiency even when evaluating a large number of candidate points. However, incorporating this approach would complicate the local convergence analysis. Therefore, we choose not to develop the partial correction framework in detail here, leaving it for future work focused on hyper-parameter-free algorithms.

Next, we establish the local quadratic convergence of Algorithm 1. A distinguishing feature of our analysis is the use of the stratification structure of the problem. By leveraging the regularity conditions that ensure favorable perturbation properties of the stratified space, we demonstrate that the proposed algorithm identifies the active stratum and converges quadratically to the solution. Importantly, these results are obtained under the weak second order condition (W-SOC, Definition 7) and the strict Robinson constraint qualification (SRCQ, Definition 2), which are much weaker than the non-degeneracy assumptions typically required in previous work to guarantee such fast convergence rates.

The proof details are provided in Appendix C. In brief, the proof is structured by decomposing the local analysis into tangential and normal directions relative to the stratum. The W-SOC and the W-SRCQ ensure a local quadratic rate along the tangent directions, as discussed in Remark 16. Furthermore, by strengthening the W-SRCQ to the SRCQ and through the specially designed algorithm, we ensure that movement along the normal directions does not result in an excessive decrease in the merit function, thereby guaranteeing the identification of $\mathcal{M}_{p,q}$.

Given the manifold $\mathcal{M} = \mathcal{M}_{p,q}$ and $\widetilde{\mathcal{M}} = G^{-1}(\mathcal{M})$, consider $z \in \widetilde{\mathcal{M}}$ and $\Delta \in \mathcal{N}_{G(z)}\mathcal{M} = (\mathcal{T}_{G(z)}\mathcal{M})^\perp$, which can be characterized by Proposition 1:

$$\mathcal{N}_{G(z)}\mathcal{M} = \left\{ H = P\widetilde{H}P^\top \mid \widetilde{H} \in \mathbb{S}^n,\ \widetilde{H}_{\alpha\alpha} = 0,\ \widetilde{H}_{\alpha\beta} = 0,\ \widetilde{H}_{\alpha\gamma} = 0,\ \widetilde{H}_{\beta\gamma} = 0,\ \widetilde{H}_{\gamma\gamma} = 0 \right\}. \tag{71}$$

Then, we have

$$F\left(z + \begin{bmatrix} 0 \\ \Delta \end{bmatrix}\right) = \begin{bmatrix} \nabla f(x) + \nabla g(x)(\lambda + \Delta) \\ -g(x) + \Pi_{\mathbb{S}^n_+}(g(x) + \lambda + \Delta) \end{bmatrix} = F(z) + \begin{bmatrix} \nabla g(x)\Delta \\ \Pi_{\mathbb{S}^n_+}(\Delta) \end{bmatrix}. \tag{72}$$

Let us define the incremental mapping as

$$J_z(\Delta) := \begin{bmatrix} \nabla g(x)\Delta \\ \Pi_{\mathbb{S}^n_+}(\Delta) \end{bmatrix}.$$

Then, for any $t \in \mathbb{R}$, we have

$$\varphi\left(z + t\begin{bmatrix} 0 \\ \Delta \end{bmatrix}\right) = \varphi(z) + 2\left\langle F(z), J_z(\Delta)\right\rangle t + \|J_z(\Delta)\|^2 t^2 \geq \left[1 - \left\langle \frac{F(z)}{\|F(z)\|}, \frac{J_z(\Delta)}{\|J_z(\Delta)\|} \right\rangle^2\right] \varphi(z), \tag{73}$$

where the inner product is defined to be 0 if either $F(z) = 0$, which implies $\varphi(z) = 0$, or $J_z(\Delta) = 0$. With this definition, it is easy to verify that (73) holds for any z.

The next two lemmas characterize the zero and the Lipschitz behavior of the incremental mapping $J_z(\Delta)$.

Lemma 7 *Let $\overline{z} = (\overline{x}, \overline{y})$ be a KKT pair of (1) with IED $(\alpha, \beta, \gamma, p, q, \overline{P}, \overline{\lambda})$ of $G(\overline{z})$. If the W-SOC and the SRCQ holds at $\overline{z}$, then 0 is the only zero of $F(z + \begin{bmatrix} 0 \\ \cdot \end{bmatrix})$ on $\mathcal{N}_{G(\overline{z})}\mathcal{M}_{p,q}$.*

Lemma 8 *For any z, J_z is positively homogeneous. If the W-SOC and the SRCQ holds at a KKT pair $\overline{z}$, there exist constants $0 < c_1 < c_2$ such that*

$$c_1\|\Delta\| \le \|J_{\overline{z}}(\Delta)\| \le c_2\|\Delta\| \quad \forall\, \Delta \in \mathcal{N}_{G(\overline{z})}\mathcal{M}_{p,q},$$

where $\mathcal{M}_{p,q}$ is given by $\overline{z}$.

The following lemma provides upper and lower bounds for the directional derivative $\varphi'(z, v)$ in terms of the LM direction, under the assumption that the W-SOC and the W-SRCQ hold.

Lemma 9 *Let $\overline{z}$ be a KKT pair of (1) with an IED $(\alpha, \beta, \gamma, p, q, \overline{P}, \overline{\lambda})$ of $G(\overline{z})$. If the W-SOC and the W-SRCQ holds at $\overline{z}$, then for $z \in \widetilde{\mathcal{M}}_{p,q}$ sufficiently close to $\overline{z}$ and $v = v^{LM}(z)$, there exists $0 < m < M$ such that*

$$m\|v\|^2 \le \|\varphi'(z, v)\| \le M\|v\|^2.$$

Next, we present an estimation result which, under the W-SOC and the SRCQ, guarantees that the value of φ will not decrease too much along the normal direction.

Proposition 10 *Let $\overline{z} = (\overline{x}, \overline{y})$ be a KKT pair of (1) with an IED $(\alpha, \beta, \gamma, p, q, \overline{P}, \overline{\lambda})$ of $G(\overline{z})$. If the W-SOC and the SRCQ hold at $\overline{z}$, then there exists a constant $c_3 > 0$ such that for any $z \in \widetilde{\mathcal{M}}_{p,q}$ sufficiently close to $\overline{z}$, $\Delta \in \mathcal{N}_{G(z)}\mathcal{M}_{p,q}$ and $t \in \mathbb{R}$,*

$$\varphi(z) \le c_3\, \varphi\left(z + t \begin{bmatrix} 0 \\ \Delta \end{bmatrix}\right). \tag{74}$$

The following proposition captures the second order behavior of φ along the LM step under the retraction curve, where the linear term is scaled due to the implicit cancellation inherent in the Gauss–Newton system. It is essential for our local analysis, as it guarantees that the line search will eventually take a unit step.

Proposition 11 *Suppose that $\overline{z} = (\overline{x}, \overline{y})$ is a KKT pair of (1) with an IED $(\alpha, \beta, \gamma, p, q, \overline{P}, \overline{\lambda})$ of $G(\overline{z})$. Then, for $z \in \widetilde{\mathcal{M}}_{p,q}$ sufficiently close to $\overline{z}$*

$$\varphi(R_z(v)) = \varphi(z) + \frac{1}{2}\varphi'(z, v) + o(\|v\|^2), \tag{75}$$

where $v = v^{LM}(z)$.

The next proposition establishes the fundamental one-step quadratic convergence rate.

Proposition 12 *Let $\overline{z}$ be a KKT pair of (1) with an IED $(\alpha, \beta, \gamma, p, q, \overline{P}, \overline{\lambda})$ of $G(\overline{z})$. If the W-SOC and the W-SRCQ holds at $\overline{z}$, then for $z \in \widetilde{\mathcal{M}}_{p,q}$ sufficiently close to $\overline{z}$, we have*

$$\big\|R_z(v) - \overline{z}\big\| = O\big(\|z - \overline{z}\|^2\big),$$

where R_z is the retraction of $\widetilde{\mathcal{M}}_{p,q}$ at z given in (66) and $v = v^{LM}(z)$ is given by (65).

With the technical machinery established in the preceding propositions and lemmas, we are now in a position to state and prove the main local convergence theorem.

Theorem 10 *Suppose z^∞ is an accumulation point of the sequence generated by Algorithm 1. If $0 < \delta < \delta(z^\infty)$ and z^∞ is a KKT pair at which the W-SOC and the SRCQ hold, then the whole sequence converges to z^∞ quadratically in the sense that*

$$\|z^{k+1} - z^\infty\| \le O(\|z^k - z^\infty\|^2).$$

In particular, the sequence eventually lies on the stratum containing z^∞.

Remark 19 The main difference between our method and conventional manifold optimization is that we do not solve a given smooth problem subject to a manifold constraint. Instead, we decompose a nonsmooth problem in Euclidean space into smooth subproblems on strata from a stratification perspective. As mentioned in the introduction, this idea is also central to the concept of partial smoothness [28], where smoothness is imposed only after restricting to an active manifold (termed "restricted smoothness" in [28, Definition 2.7(i)]). Partial smoothness, however, also requires a "normal sharpness" property [28, Definition 2.7(iii)]. For the merit function φ given by (55), this property fails, so the partial smoothness theory is not directly applicable here.

Along a different line, Deng et al. [12] proposed a manifold based primal dual semismooth Newton method for (1). Their approach also combines the correction idea of [19] to drive the sequence toward the active manifold.

Superlinear convergence was established (see [12, Theorem 10]) under an error bound condition together with an assumption that the iterates eventually lie on the active manifold. In contrast, the present approach does not require the error bound to hold over the entire ambient space, nor does it rely on the prior existence of iterates on the active stratum. Instead, by investigating the stratification, we have distilled an equivalent condition required for the convergence of Newton's method applied to the manifold-restricted subproblem, which is shown to be significantly weaker than classical regularity conditions. Furthermore, by exploiting the variational properties induced by the stratification, the proposed algorithm, through a specific design, drives the iterates toward the active stratum and ultimately achieves manifold identification.

5 Conclusion

This paper develops a stratification-based viewpoint for nonlinear semidefinite programming. By viewing $\mathbb{S}^n$ through the index stratification and lifting it to the primal–dual space, the nonsmooth KKT system is endowed with a hidden smooth geometry: the projector onto $\mathbb{S}^n_+$ is C^∞ on each stratum, and the KKT residual admits an explicit differential once restricted to a fixed stratum. This perspective cleanly separates *variational analysis on a stratum*, where classical smooth tools apply, from *analysis across strata*, where transitions are dictated by the adjacency relations of the stratification. Within this framework, weak and verifiable problem-level regularity conditions are linked to problem-level solution-level regularity conditions in a stratum-restricted sense, and classical strong-form regularity conditions are shown to correspond to a local uniform validity of these stratum-restricted properties near a solution. The transversality interpretation clarifies the geometry underlying the W-SRCQ and yields stability and genericity results that parallel and, in several respects, extend the classical theory. Stratification theory also leads naturally to algorithmic design. A globalized stratified Gauss–Newton scheme is proposed, equipped with normal steps and a correction operator, so as to ensure descent for a least-squares merit function while allowing controlled movement between strata. Under suitable assumptions, the method converges globally to directional stationary points and enjoys local quadratic convergence to KKT pairs, with eventual identification of the active stratum.

Beyond its specific implications for the KKT system, the proposed stratification framework should be viewed as a starting point rather than an endpoint. It may provide a geometric lens for NLSDP and, more broadly, nonpolyhedral matrix optimization, by making explicit the stratified smooth structure that is often hidden behind nonsmooth formulations. On the theoretical side, an important direction is to revisit other central notions in perturbation analysis, including the robust isolated calmness, the Aubin property, and related other stability concepts. Through this stratified viewpoint, with the expectation that stratified smoothness may lead to sharper characterizations. On the algorithmic side, stratification also opens several new research directions. It may support the systematic design of efficient methods for degenerate regimes in which classical assumptions are overly restrictive by combining stratum-restricted regularity with principled stratum-identification mechanisms to obtain fast local rates under weaker hypotheses. It also motivates extending the present analysis beyond Newton-type schemes to other NLSDP solvers, notably augmented Lagrangian methods and related approaches. Finally, it is of clear practical interest to eliminate the dependence on the eigenvalue-threshold hyperparameter, for instance, by developing partial correction strategies that retain both robust globalization and fast local rate theory.

References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization algorithms on matrix manifolds. Princeton University Press (2008)
2. Arnold, E., Gusein-Zade, S.M., Varchenko, A.N.: Singularities of differentiable maps, volume 2: monodromy and asymptotics of integrals. Springer Science & Business Media (2012)
3. Bhatia, R.: Matrix analysis, vol. 169. Springer Science & Business Media (2013)
4. Bonnans, J.F., Cominetti, R., Shapiro, A.: Sensitivity analysis of optimization problems under second order regular constraints. Mathematics of Operations Research **23**(4), 806–831 (1998)
5. Bonnans, J.F., Cominetti, R., Shapiro, A.: Second order optimality conditions based on parabolic second order tangent sets. SIAM Journal on Optimization **9**(2), 466–492 (1999)
6. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer Verlag (2000)
7. Bonnans, J.F., Sulem, A.: Pseudopower expansion of solutions of generalized equations and constrained optimization problems. Mathematical Programming **70**(1-3), 123 – 148 (1995). DOI 10.1007/bf01585932
8. Boumal, N., Voroninski, V., Bandeira, A.S.: Deterministic guarantees for Burer-Monteiro factorizations of smooth semidefinite programs. Communications on Pure and Applied Mathematics **73**(3), 581–608 (2020)
9. Burer, S., Monteiro, R.D.: A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. Mathematical programming **95**(2), 329–357 (2003)
10. Burer, S., Monteiro, R.D.: Local minima and convergence in low-rank semidefinite programming. Mathematical programming **103**(3), 427–444 (2005)
11. Conway, J.B.: A course in functional analysis, vol. 96. Springer Science & Business Media (1994)

12. Deng, Z., Hu, J., Deng, K., Wen, Z.: An efficient primal dual semismooth newton method for semidefinite programming. arXiv preprint arXiv:2504.14333 (2025)
13. Ding, C., Sun, D., Zhang, L.: Characterization of the robust isolated calmness for a class of conic programming problems. SIAM Journal on Optimization **27**(1), 67–90 (2017)
14. Dontchev, A.L., Rockafellar, R.T.: Characterizations of strong regularity for variational inequalities over polyhedral convex sets. SIAM Journal on Optimization **6**(4), 1087 – 1105 (1996). DOI 10.1137/s1052623495284029
15. Dontchev, A.L., Rockafellar, R.T.: Implicit functions and solution mappings, vol. 543. Springer (2009)
16. Van den Dries, L.: Tame topology and o-minimal structures, vol. 248. Cambridge university press (1998)
17. Drusvyatskiy, D., Lewis, A.S.: Optimality, identifiability, and sensitivity. Mathematical Programming **147**(1), 467–498 (2014)
18. Facchinei, F., Pang, J.S.: Finite-dimensional variational inequalities and complementarity problems. Springer (2003)
19. Feng, F., Ding, C., Li, X.: A quadratically convergent semismooth newton method for nonlinear semidefinite programming without generalized jacobian regularity. Mathematical Programming pp. 1–41 (2025)
20. Gao, B., Absil, P.A.: A riemannian rank-adaptive method for low-rank matrix completion. Computational Optimization and Applications **81**(1), 67–90 (2022)
21. Goresky, M., MacPherson, R.: Stratified morse theory. In: Stratified Morse Theory, pp. 3–22. Springer (1988)
22. Helmke, U., Shayman, M.A.: Critical points of matrix least squares distance functions. Linear Algebra and its Applications **215**, 1–19 (1995). DOI https://doi.org/10.1016/0024-3795(93)00070-G. URL `https://www.sciencedirect.com/science/article/pii/002437959300070G`
23. Hirsch, M.W.: Differential topology, vol. 33. Springer Science & Business Media (2012)
24. Ioffe, A.D.: An invitation to tame optimization. SIAM Journal on Optimization **19**(4), 1894–1917 (2009)
25. Jongen, H.T., Möbert, T., Rückmann, J., Tammer, K.: On inertia and schur complement in optimization. Linear Algebra and its Applications **95**, 97–109 (1987)
26. Jost, J.: Riemannian geometry and geometric analysis. Springer (2005)
27. Lee, J.: Introduction to Smooth Manifolds, vol. 218. Springer Science & Business Media (2012)
28. Lewis, A.S.: Active sets, nonsmoothness, and sensitivity. SIAM Journal on Optimization **13**(3), 702–725 (2002)
29. Ma, J., Sun, D.: The Aubin property for generalized equations over C^2-cone reducible sets. arXiv preprint arXiv:2509.14194 (2025)
30. Mather, J.: Notes on topological stability, vol. 6. Harvard University Cambridge (1970)
31. Molina, J.L.C., Lê, D.T., Seade, J.: Handbook of Geometry and Topology of Singularities I. Springer (2020)
32. Olikier, G., Gallivan, K.A., Absil, P.A.: First-order optimization on stratified sets. ArXiv **abs/2303.16040** (2023). URL `https://api.semanticscholar.org/CorpusID:257771438`
33. Qi, L.: Convergence analysis of some algorithms for solving nonsmooth equations. Mathematics of operations research **18**(1), 227–244 (1993)
34. Qi, L., Sun, J.: A nonsmooth version of newton's method. Mathematical programming **58**(1), 353–367 (1993)
35. Robinson, S.M.: Strongly regular generalized equations. Mathematics of Operations Research **5**, 43–62 (1980)
36. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis, *Springer Berlin Heidelberg*, vol. 317. Springer Berlin Heidelberg (1998). DOI 10.1007/978-3-642-02431-3. URL `http://link.springer.com/10.1007/978-3-642-02431-3`
37. Simon, B.: Operator theory, vol. 4. American Mathematical Soc. (2015)
38. Smith, S.T.: Optimization techniques on riemannian manifolds. arXiv preprint arXiv:1407.5965 (2014)
39. Sun, D.: The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. Mathematics of Operations Research **31**(4), 761–776 (2006)
40. Sun, D., Sun, J.: Semismooth matrix-valued functions. Mathematics of Operations Research **27**(1), 150–169 (2002)
41. Thom, R.: Ensembles et morphismes stratifiés (1969)
42. Vandereycken, B.: Low-rank matrix completion by Riemannian optimization. SIAM Journal on Optimization **23**(2), 1214–1236 (2013)
43. Whitney, H.: Local properties of analytic varieties. Differential and Combinatorial Topology pp. 205–244 (1965)
44. Whitney, H.: Tangents to an analytic variety. The Annals of Mathematics **81**(3), 496 (1965)
45. Yamashita, N., Fukushima, M.: On the rate of convergence of the Levenberg–Marquardt method. Computing **67**(3), 239–252 (2001). DOI 10.1007/s006070170004
46. Zhou, G., Huang, W., Gallivan, K.A., Van Dooren, P., Absil, P.A.: A riemannian rank-adaptive method for low-rank optimization. Neurocomputing **192**, 72–80 (2016)

A Smoothness of $\Pi_{\mathbb{S}^n_+}$ on the fixed-index manifold

In this appendix, we prove the C^∞ smoothness of $\Pi_{\mathbb{S}^n_+}$ on the manifold $\mathcal{M}_r := \{A \in \mathbb{S}^n \mid \text{rank}A = r\}$ and compute its differential. In fact, for any $r = 1, ..., n$, the symmetric matrices with rank r have $r+1$ connected components, $\mathcal{M}_r = \cup_{p=0}^r \mathcal{M}_{p,r-p}$. So we only need to prove the C^∞ smoothness of $\Pi_{\mathbb{S}^n_+}$ and compute its differential on the manifold $\mathcal{M}_{p,q}$ for any $p, q \in \mathbb{Z}_+$.

The C^∞ smoothness of $\Pi_{\mathbb{S}^n_+}$ on the manifold $\mathcal{M}_{p,q}$ implies that $W_1(z)$ and $W_2(z)$ defined in Proposition 8 are both smooth on $\widetilde{\mathcal{M}}_{p,q}$, which is Lemma C.4. Along with the above property, we will also prove that the retraction we defined in Section 4.1.2 is also smooth on the manifold $\widetilde{\mathcal{M}}_{p,q}$ for any p, q. As before, we will drop the indices p, q and just write $\mathcal{M}$ and $\widetilde{\mathcal{M}}$ for simplicity since the precise p, q is not important for us. The proof is based on simple and direct application of functional calculus, which might not be familiar to some applied mathematicians but is a classical subject in operator algebra.

Roughly speaking, in complex analysis, the Cauchy integral formula allows us to recover an analytic function $f(z)$ using a contour integral:

$$f(z) = \frac{1}{2\pi i} \int_\Gamma \frac{f(\zeta)}{\zeta - z} d\zeta, \tag{76}$$

where Γ encloses z. This formula not only reconstructs f but also permits differentiation under the integral, implying f is analytic (hence infinitely differentiable). For a symmetric matrix $A \in \mathbb{S}^n$, a similar process called *analytic functional calculus* defines $f(A)$ for an analytic f via a similar contour integral:

$$f(A) = \frac{1}{2\pi i} \int_\Gamma f(z)(zI - A)^{-1} dz. \tag{77}$$

Here, Γ encloses the spectrum (i.e. its eigenvalues) of A and $(zI - A)^{-1}$ is called the resolvent of A. This mirrors the Cauchy formula. For a serious treatment, we will recommend [11, Chapter VII], which is a classical textbook on advanced functional analysis, or [37, Sections 2.2 & 2.3], which covers a detailed discussion and some historical materials. The following theorem is taken from [11, Chapter VII].

Theorem A.1. *For any f analytic in an open set Ω containing* $\mathrm{Spec}(A)$ *and $\Gamma = \partial\Omega$ encloses* $\mathrm{Spec}(A)$*,*

$$f(A) = \frac{1}{2\pi i} \int_\Gamma f(z)(zI - A)^{-1} dz$$

depends analytically on $A \in \mathbb{S}^n$ (i.e., it is infinitely differentiable and locally expressible as a convergent power series).

We take some particular functions to prove our results. Since $A \in \mathcal{M}$ has fixed $\alpha(A), \beta(A)$ and $\gamma(A)$, we can find three disjoint open disks $D_i, i = 1, 2, 3$ in the complex number field $\mathbb{C}$, each of which contains the positive and negative parts of eigenvalues of A and 0. And we denote $D = D_1 \cup D_2 \cup D_3$. When A varies slightly on $\mathcal{M}$, the positive, negative and zero parts of $\lambda(A)$ will stay in D_1, D_2 and D_3 respectively. Therefore, taking analytic functions

$$f_+(z) := \begin{cases} z, z \in D_1 \\ 0, z \in D_2 \cup D_3, \end{cases} \qquad f_\dagger(z) := \begin{cases} \dfrac{1}{z}, z \in D_1 \cup D_2 \\ 0, z \in D_3, \end{cases} \qquad \text{and} \quad p(z) := \begin{cases} 1, z \in D_1 \cup D_2 \\ 0, z \in D_3 \end{cases}$$

on D, we can get operators $f_+(A)$ and $f_\dagger(A)$ analytic with respect to A by the above theorem. Now we can prove Theorem 1.

Proof (of Theorem 1) For any $A \in \mathcal{M} \subseteq \mathbb{S}^n$, take the exponential local chart $\exp_A : \mathcal{U} \subseteq \mathcal{T}_A\mathcal{M} \to \mathcal{U}' \subseteq \mathcal{M}$ around A in $\mathcal{M}$. Then by the analyticity of $f_+(A)$ with respect to A, we have a smooth mapping:

$$\begin{array}{ccccc} \mathcal{U} & \longrightarrow & \mathcal{U}' & \longrightarrow & \mathbb{S}^n \\ v & \longmapsto & \exp_A(v) & \longmapsto & f_+(\exp_A(v)) \end{array} \tag{78}$$

For any $\widetilde{A} \in \mathcal{M}$, by direct computation, we will have $\Pi_{\mathbb{S}^n_+}(\widetilde{A}) = f_+(\widetilde{A})$. Therefore, the map

$$v \mapsto f_+(\exp_A(v)) = \Pi_{\mathbb{S}^n_+}(\exp_A(v))$$

is smooth, which by definition of smoothness means $\Pi_{\mathbb{S}^n_+} : \mathcal{M} \to \mathbb{S}^n$ is smooth.

The above is the first half of Theorem 1 and now we compute its differential. For any $A \in \mathcal{M}$ and $H \in \mathcal{T}_A\mathcal{M}$, as $t \to 0$, we have

$$\Pi'_{\mathbb{S}^n_+}(A; H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \Xi_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\alpha\beta}^\top & \Pi_{\mathbb{S}^{|\beta|}_+}(0) & 0 \\ \widetilde{H}_{\alpha\gamma}^\top \circ \Xi_{\alpha\gamma}^\top & 0 & 0 \end{bmatrix} P^\top = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \Xi_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\alpha\beta}^\top & 0 & 0 \\ \widetilde{H}_{\alpha\gamma}^\top \circ \Xi_{\alpha\gamma}^\top & 0 & 0 \end{bmatrix} P^\top.$$

The second equality comes from $\widetilde{H}_{\beta\beta} = 0$ by Proposition 1. The varying term in the RHS of (5) will vanish. By the definition of differential, we have

$$d(\Pi_{\mathbb{S}^n_+})_A(H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \Xi_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\alpha\beta}^\top & 0 & 0 \\ \widetilde{H}_{\alpha\gamma}^\top \circ \Xi_{\alpha\gamma}^\top & 0 & 0 \end{bmatrix} P^\top.$$

This completes the proof. □

Finally, we prove the smoothness of $\pi^2_{z,\mathbb{S}^n}$ (defined in (59)) with respect to z, which is used in Lemma C.4.

Theorem A.2. *For $z \in \widetilde{\mathcal{M}}_{p,q}$, as above we define a piecewise constant (hence analytic) function*

$$p_0(w) := \begin{cases} 0, w \in D_3 \\ 1, w \in D_1 \cup D_2, \end{cases}$$

where D_1, D_2 and D_3 contains the positive and negative parts of eigenvalues of $G(z)$ and 0 respectively, then for any $H \in \mathbb{S}^n$ there holds

$$\pi^2_{z,\mathbb{S}^n}(H) = p_0(G(z))Hp_0(G(z)),$$

where $\pi^2_{z,\mathbb{S}^n} : \mathbb{S}^n \to \mathcal{N}_{G(z)}\mathcal{M}_{p,q}$ is the projection defined according to Lemma 4 .

In particular, the smoothness of $p_0(G(z))$ with respect to z implies the smoothness of $\pi^2_{z,\mathbb{S}^n}(H)$.

Proof We only need to check that $\pi^2_{z,\mathbb{S}^n}(H) = p_0(G(z))Hp_0(G(z))$. Suppose z has an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$. Denote $P = [P_\alpha\, P_\beta\, P_\gamma]$. Then $P_\alpha, P_\beta, P_\gamma$ are an orthonormal basis of positive, zero, and negative eigen spaces of $G(z)$, respectively. So $P_\beta P_\beta^\top = p_0(G(z))$ since they are both the projection matrix onto the zero eigen space of $G(z)$.

By definition, we have

$$\pi^2_{z,\mathbb{S}^n}(H) = P\pi_{\beta\beta}(P^\top HP)P^\top,$$

where $\pi_{\beta\beta}$ is defined to be

$$\pi_{\beta\beta}\left(\begin{bmatrix} X_{\alpha\alpha} & X_{\alpha\beta} & X_{\alpha\gamma} \\ X_{\beta\alpha} & X_{\beta\beta} & X_{\beta\gamma} \\ X_{\gamma\alpha} & X_{\gamma\beta} & X_{\gamma\gamma} \end{bmatrix}\right) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & X_{\beta\beta} & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0\,0\,0 \\ 0\,I\,0 \\ 0\,0\,0 \end{bmatrix} X \begin{bmatrix} 0\,0\,0 \\ 0\,I\,0 \\ 0\,0\,0 \end{bmatrix}.$$

Therefore, we have

$$\pi^2_{z,\mathbb{S}^n}(H) = P\pi_{\beta\beta}(P^\top HP)P^\top = P\begin{bmatrix} 0\,0\,0 \\ 0\,I\,0 \\ 0\,0\,0 \end{bmatrix}(P^\top HP)\begin{bmatrix} 0\,0\,0 \\ 0\,I\,0 \\ 0\,0\,0 \end{bmatrix}P^\top = (P_\beta P_\beta^\top)H(P_\beta P_\beta^\top) = p_0(G(z))Hp_0(G(z)),$$

which completes the proof. □

B Details about transversality

In this part, we will give the rigorous proofs of results in Section 3.1.2. First, we recall the rigorous definition of "generic" and list some theorems we will use.

Definition B.1. [23, Chapter 3.1] A subset of a space X is called *residual* if it contains the intersection of countably many dense open sets. And we say some property parametrized by $v \in \mathcal{V}$ holds *genericly* if it holds for a residual subset in $\mathcal{V}$.

The following theorem describes the stability and genericity of transversality.

Theorem B.1. *[23, Theorem 3.2.1] Let $\mathcal{M}, \mathcal{N}$ be manifolds and $\mathcal{N}' \subseteq \mathcal{N}$ a submanifold. Define*

$$\pitchfork_{\mathcal{L}}(\mathcal{M}, \mathcal{N}; \mathcal{N}') = \{f \in C^\infty(\mathcal{M}, \mathcal{N}) : f \pitchfork_{\mathcal{L}} \mathcal{N}'\} \text{ and } \pitchfork(\mathcal{M}, \mathcal{N}; \mathcal{N}') = \pitchfork_{\mathcal{M}}(\mathcal{M}, \mathcal{N}; \mathcal{N}').$$

Then, we have

(a) $\pitchfork(\mathcal{M}, \mathcal{N}; \mathcal{N}')$ is residual (and therefore dense) in $C^\infty(\mathcal{M}, \mathcal{N})$ for both the strong and weak topologies;
(b) suppose $\mathcal{N}'$ is closed in $\mathcal{N}$. If $\mathcal{L} \subseteq \mathcal{M}$ is closed [resp. compact], then $\pitchfork_{\mathcal{L}}(\mathcal{M}, \mathcal{N}; \mathcal{N}')$ is dense and open in $C^\infty(\mathcal{M}, \mathcal{N})$ with strong topology [resp. with weak topology].

The following theorem is about parametrized transversality, which coincides with our construction. Without any surprise, we will apply it to our proof.

Theorem B.2. *[23, Theorem 3.2.7] Let $\mathcal{V}, \mathcal{M}, \mathcal{N}$ be smooth manifolds without boundary and $\mathcal{N}' \subseteq \mathcal{N}$ a smooth submanifold. Let $F : \mathcal{V} \to C^\infty(\mathcal{M}, \mathcal{N})$ satisfy the following conditions:*

(a) the evaluation map $F^{\text{ev}} : \mathcal{V} \times \mathcal{M} \to \mathcal{N}, (v, x) \mapsto F_v(x)$, is C^∞;
(b) F^{ev} is transverse to $\mathcal{N}'$.

Then the set

$$\pitchfork(F;\mathcal{N}') := \{v \in \mathcal{V} : F_v \pitchfork \mathcal{N}'\}$$

is residual and therefore dense. If $\mathcal{N}'$ is closed in $\mathcal{N}$ and F is continuous for the strong topology on $C^\infty(\mathcal{M},\mathcal{N})$ then $\pitchfork(F,\mathcal{N}')$ is also open.

Secondly, we introduce a simplified notion of tubular neighborhood, which is illustrated by its name, a neighborhood looks like a tube.

Definition B.2. [27, Chapter 6] Let $\mathbb{Y}$ be a Euclidean space and $\mathcal{S} \subset \mathbb{Y}$ be an embedded submanifold. The *normal bundle* $\mathcal{NS}$ of $\mathcal{S}$ in $\mathbb{Y}$ is the vector bundle over $\mathcal{S}$ defined fiberwise by:

$$\mathcal{N}_x\mathcal{S} = (\mathcal{T}_x\mathcal{S})^\perp \text{ for each } x \in \mathcal{S},$$

where the orthogonal complement is given by the standard inner product of $\mathbb{Y}$.

In our case, $\mathbb{Y}$ is $\mathbb{S}^n$ and $\mathcal{S}$ is $\mathcal{M}_{p,q}$ for some p,q. Since from Proposition 1 we know $\mathcal{T}_A\mathcal{M}_{p,q} = \left\{H \in \mathbb{S}^n \mid (P^\top HP)_{\beta\beta} = 0\right\}$ for some IED $(\alpha,\beta,\gamma,p,q,P,\lambda)$ of A, it is easy to see

$$\mathcal{N}_A\mathcal{M}_{p,q} = \left\{H \in \mathbb{S}^n \mid \text{ only } (P^\top HP)_{\beta\beta} \text{ can be nonzero } \right\}.$$

Therefore, $\mathcal{NM}_{p,q} = \cup_{A\in\mathcal{M}_{p,q}}\{A\} \times \mathcal{N}_A\mathcal{M}_{p,q}$ with a compatible smooth structure.

It can be shown that $E : \mathbb{Y} \times \mathbb{Y} \to \mathbb{Y}, (y_1,y_2) \mapsto y_1 + y_2$ is a diffeomorphism when restricted to a small neighborhood of $\mathcal{S}$ in $\mathcal{NS}$. We recommend [27, Chapter 6] for more details.

Definition B.3. [27, Chapter 6] A *tubular neighborhood* of $\mathcal{S}$ in $\mathbb{Y}$ is an open neighborhood $\mathcal{DS} \subseteq \mathbb{Y}$ of $\mathcal{S}$, which is the image under the diffeomorphism $E : \mathcal{W} \to \mathcal{DS}$, where $\mathcal{W}$ is an open neighborhood of the $\mathcal{S}$ in $\mathcal{NS}$.

With the above preparations, we can give the rigorous construction of $\mathcal{DN}_z$ step by step. First, we define a family of submanifolds

$$\mathcal{N}_z := \mathcal{M}_{p,0} + \Pi_{\mathbb{S}^n_-}(G(z)) - y \tag{79}$$

parametrized by $z = (x,y)$, where p is the positive inertia index of $G(z)$. Moreover, when z varies on some stratum $\widetilde{\mathcal{M}}_{p,q}$, $\{\mathcal{N}_z\}_{z\in\widetilde{\mathcal{M}}_{p,q}}$ forms a smooth family and the following lemma shows that for any $z \in \widetilde{\mathcal{M}}_{p,q}$, there is a small open set $\mathcal{U}$ in $\mathcal{M}_{p,0}$ such that $\mathcal{U} + \{\Pi_{\mathbb{S}^n_-}(G(z))\} - y$ is a neighborhood of $g(x)$ in $\mathcal{N}_z$ and embedded in $\mathcal{M}_{p,q} - y$. Moreover, shrinking $\mathcal{U}$ (to be a compact one) if necessary, we have $\mathcal{U} + \{\Pi_{\mathbb{S}^n_-}(G(z')) - y'\}$ are all embedded in $\mathcal{M}_{p,q}$ for any $z' = (x',y') \in \widetilde{\mathcal{M}}_{p,q}$ near $z = (x,y) \in \widetilde{\mathcal{M}}_{p,q}$. Since transversality is local, we can always shrink the objects. Therefore we may just assume that $\mathcal{N}_z$ are embedded in $\mathcal{M}_{p,q} - y$, otherwise we just define $\mathcal{N}_z$ to be $\mathcal{U} + \{\Pi_{\mathbb{S}^n_-}(G(z)) - y\}$. Also, we may assume that there exists a neighborhood $\widetilde{\mathcal{U}}$ of z in $\widetilde{\mathcal{M}}_{p,q}$ such that $\cup_{z'\in\widetilde{\mathcal{U}}}(\mathcal{N}_{z'} \times \{z'\})$ is a smooth submanifold in $\mathcal{M}_{p,q} \times \widetilde{\mathcal{M}}_{p,q}$, or to be illustrating, $\mathcal{N}_{z'}$ varies smoothly when z' varies smoothly in $\widetilde{\mathcal{M}}_{p,q}$ near z.

Lemma B.1. *Let $z \in \widetilde{\mathcal{M}}_{p,q}$ and set $A := G(z) \in \mathcal{M}_{p,q}$. Write*

$$A_+ := \Pi_{\mathbb{S}^n_+}(A) \in \mathcal{M}_{p,0}, \quad A_- := \Pi_{\mathbb{S}^n_-}(A) \in \mathcal{M}_{0,q},$$

so that $A = A_+ + A_-$. Then there exists a neighborhood $\mathcal{U}$ of A in $\mathcal{M}_{p,0} + A_-$ such that $\mathcal{U}$ is smoothly embedded in $\mathcal{M}_{p,q}$. Therefore, there exists a neighborhood $\mathcal{U}$ of $g(x) = A - y$ in $\mathcal{M}_{p,0} + A_- - y$ such that $\mathcal{U}$ is smoothly embedded in $\mathcal{M}_{p,q} - y$, which by definition means $\mathcal{N}_z$ is locally embedded in $\mathcal{M}_{p,q} - y$ near $g(x)$.

Proof By the eigen-decomposition of A in (4), we know that

$$P_\alpha^T A_+ P_\alpha \succ 0 \quad \text{on} \quad \mathbb{R}^p, \tag{80}$$

$$P_\alpha^T A_- P_\alpha = 0 \quad \text{on} \quad \mathbb{R}^p, \quad P_\beta^T A_- P_\beta = 0 \quad \text{on} \quad \mathbb{R}^{n-p-q}, \text{ and } P_\gamma^T A_- P_\gamma \prec 0 \quad \text{on} \quad \mathbb{R}^q. \tag{81}$$

By (80) and the continuity of eigenvalues, there is a neighborhood U of A_+ that for any $A'_+ \in \mathcal{M}_{p,0} \cap U$

$$P_\alpha^T A'_+ P_\alpha \succ 0 \quad \text{on} \quad \mathbb{R}^p. \tag{82}$$

Since $A'_+ \in \mathcal{M}_{p,0}$, (82) gives the p positive eigenvalues of A'_+ and $P_\alpha, P_\beta, P_\gamma$ are orthogonal, we have

$$P_\beta^T A'_+ P_\beta = 0 \quad \text{on} \quad \mathbb{R}^{n-p-q} \quad \text{and} \quad P_\gamma^T A'_+ P_\gamma = 0 \quad \text{on} \quad \mathbb{R}^q. \tag{83}$$

Put them together, we get

$$P_\alpha^T (A'_+ + A_-) P_\alpha \succ 0 \quad \text{on} \quad \mathbb{R}^p \tag{84}$$

$$P_\beta^T (A'_+ + A_-) P_\beta = 0 \quad \text{on} \quad \mathbb{R}^{n-p-q} \tag{85}$$

$$P_\gamma^T (A'_+ + A_-) P_\gamma \prec 0 \quad \text{on} \quad \mathbb{R}^q, \tag{86}$$

which shows $A'_+ + A_- \in \mathcal{M}_{p,q}$. The proof is then completed. □

Now consider the normal bundle

$$\mathcal{N}(\mathcal{M}_{p,q}-y):=\cup_{A\in\mathcal{M}_{p,q}}\mathcal{N}_A(\mathcal{M}_{p,q}-y)\subseteq\mathbb{S}^n\times\mathbb{S}^n$$

of $\mathcal{M}_{p,q}-y$ in $\mathbb{S}^n$. And we can restrict the normal bundle of $\mathcal{M}_{p,q}-y$ to its submanifolds $\mathcal{N}_z$ and get a smooth family of vector bundles $\{\mathcal{N}(\mathcal{M}_{p,q}-y)|_{\mathcal{N}_z}\}_{z\in\widetilde{\mathcal{M}}_{p,q}}$. On the other hand, by [27, Chapter 6.4-Tubular neighborhood], we know the map

$$E:\mathbb{S}^n\times\mathbb{S}^n\to\mathbb{S}^n \tag{87}$$

$$(A,H)\mapsto A+H \tag{88}$$

is a diffeomorphism from a neighborhood of $\mathcal{M}_{p,q}$ in $\mathcal{NM}_{p,q}$ to a tubular neighborhood $\mathcal{DM}_z$ of $\mathcal{M}_{p,q}-y$ in $\mathbb{S}^n$. Finally, we can define a smooth family of smooth submanifolds with the above two objects. We define a family of submanifolds in $\mathbb{S}^n$:

$$\{\mathcal{DN}_z:=(E(\mathcal{N}(\mathcal{M}_{p,q}-y)|_{\mathcal{N}_z})\cap\mathcal{DM}_z)\}_{z\in\widetilde{\mathcal{M}}_{p,q}}.$$

After defining $\mathcal{DN}_z$, we can prove Theorem 3 now.

Theorem 3 (restated). *For the optimization problem* (1)*, for any $z=(x,y)\in\mathbb{X}\times\mathbb{S}^n$ with an IED $(\alpha,\beta,\gamma,p,q,P,\lambda)$ of $G(z)$ and the corresponding manifolds $\mathcal{M}_{p,q}$, the following are equivalent:*

(i) $g\pitchfork_x\mathcal{DN}_z$ in $\mathbb{S}^n$;
(ii) the W-SRCQ holds at $z=(x,y)$.

Proof Let us start from (i), which is just saying that

$$dg_x(\mathbb{X})+\mathcal{T}_{g(x)}(\mathcal{DN}_z)=\mathbb{S}^n. \tag{89}$$

With an IED $(\alpha,\beta,\gamma,p,q,P,\lambda)$, we can express the tangent space $\mathcal{T}_{g(x)}(\mathcal{DN}_z)$ in $\mathbb{S}^n$ as

$$\begin{aligned}\mathcal{T}_{g(x)}(\mathcal{DN}_z)&=\mathcal{T}_{G(z)}(E(\mathcal{NM}_{p,q}|_{\mathcal{N}_z}))&(90)\\&=\mathcal{T}_{G(z)}\mathcal{N}_z+\mathcal{N}_{G(z)}\mathcal{M}_{p,q}&(91)\\&=\left\{PBP^\top\mid B\in\mathbb{S}^n,\ B_{\beta\beta}=0,\ B_{\beta\gamma}=0,\ B_{\gamma\gamma}=0\right\}&(92)\\&\quad+\left\{PBP^\top\mid B\in\mathbb{S}^n\text{ and only }B_{\beta\beta}\text{ can be nonzero}\right\}&(93)\\&=\left\{PBP^\top\mid B\in\mathbb{S}^n,\ B_{\beta\gamma}=0,\ B_{\gamma\gamma}=0\right\},&(94)\end{aligned}$$

where the second equality comes from the openness of $\mathcal{DM}_{p,q}$, the third term is computed directly by the definition of E, which is just an vector addition. Therefore, we can see (i) is equivalent to

$$dg_x(\mathbb{X})+\left\{PBP^\top\mid B\in\mathbb{S}^n,\ B_{\beta\gamma}=0,\ B_{\gamma\gamma}=0\right\}=\mathbb{S}^n, \tag{95}$$

which is nothing but (ii). □

We are now able to present the proof of Theorem 4.

Theorem 4 (restated). *If W-SRCQ holds at some $\overline{z}=(\overline{x},\overline{y})\in\widetilde{\mathcal{M}}_{p,q}$, then there will exist a neighborhood $\widetilde{\mathcal{U}}\subseteq\widetilde{\mathcal{M}}_{p,q}$ of $\overline{z}$ such that W-SRCQ holds for any $z\in\widetilde{\mathcal{U}}$.*

Proof We prove this theorem through a purely differential topology approach. Before the rigorous proof, we give some explanations. Since Theorem 3 tells us W-SRCQ holding at $z=(x,y)\in\widetilde{\mathcal{M}}_{p,q}$ is equivalent to $g\pitchfork_x\mathcal{DN}_z$ in $\mathbb{S}^n$, then the stability of W-SRCQ near $\overline{z}\in\widetilde{\mathcal{M}}_{p,q}$ is equivalent to the stability of this transversality, i.e. $g\pitchfork_x\mathcal{DN}_z$ in $\mathbb{S}^n$ holds for all $z\in\widetilde{\mathcal{M}}_{p,q}$ near $\overline{z}$. So we have finished the proof if you believe in the stability of this transversality, which is slightly different from the classical notion. The classical stability of transversality [23, Theorem 3.2.7] is for the case that the intersecting manifold is fixed. However, in our case, $\mathcal{DN}_{z'}$ is parametrized, which seems different. Fortunately, we can transform our case into the classical case. The following detailed proof is for readers who might be interested, which is just a purely technical process – transforming the varying of $\mathcal{DN}_{z'}$ into the varying of maps and applying the classical results.

Just like what we mentioned before Lemma B.1, by taking a more careful shrinking, we may assume that for any z' in a small neighborhood $\widetilde{\mathcal{U}}$ of $\overline{z}$ in $\widetilde{\mathcal{M}}_{p,q}$, we have $\mathcal{N}_{z'}=\mathcal{U}+\{\mathit{\Pi}_{\mathbb{S}^n_-}(G(z'))-y'\}$, which is embedded in $\mathcal{M}_{p,q}-y'$, and $\mathcal{DN}_{z'}=E\big(\mathcal{N}(\mathcal{M}_{p,q}-y')|_{\mathcal{N}_{z'}}\big)$. Denote the embeddings of $\mathcal{N}_{z'}$ by

$$i_{z'}:\mathcal{U}\to\mathcal{N}_{z'}\subseteq\mathcal{M}_{p,q}-y' \tag{96}$$

$$A\mapsto A+\mathit{\Pi}_{\mathbb{S}^-_n}(G(z'))-y'. \tag{97}$$

When z' varies on $\widetilde{\mathcal{M}}_{p,q}$, $i_{z'}$ forms a smooth family of embeddings and induce a smooth family of isomorphisms of pullback bundles

$$\left\{\psi_{z'} : \mathcal{N}(\mathcal{M}_{p,q} - \overline{y})|_{\mathcal{N}_{\overline{z}}} \xrightarrow{\cong} \mathcal{N}(\mathcal{M}_{p,q} - y')|_{\mathcal{N}_{z'}}\right\}_{z' \in \widetilde{\mathcal{U}}},$$

satisfying

(i) $\psi_{\overline{z}} = id$;
(ii) $\psi_{z'}|_{\mathcal{N}_{g(\overline{x})}(\mathcal{M}_{p,q} - \overline{y})}$ is an isomorphism from $\mathcal{N}_{g(\overline{x})}(\mathcal{M}_{p,q} - \overline{y})$ to $\mathcal{N}_{g(x')}(\mathcal{M}_{p,q} - y')$.

Therefore, $\mathcal{DN}_{z'}$ is just the image of the embeddings $j_{z'} = E \circ \psi_{z'} : \mathcal{N}(\mathcal{M}_{p,q} - \overline{y})|_{\mathcal{N}_{\overline{z}}} \to \mathbb{S}^n$, which is a smooth family of maps parametrized by $z' \in \widetilde{\mathcal{U}}$.

Now consider the parametrized map

$$\Phi_{z'} := (g, j_{z'}) : \mathbb{X} \times \mathcal{N}(\mathcal{M}_{p,q} - \overline{y})|_{\mathcal{N}_{\overline{z}}} \to \mathbb{S}^n \times \mathbb{S}^n \tag{98}$$

$$(x, \zeta) \mapsto (g(x), j_{z'}(\zeta)). \tag{99}$$

Noting that $\mathcal{T}_{j_{z'}(\zeta)}\mathcal{DN}_{z'} = d(j_{z'})_\zeta(\mathcal{T}_\zeta \mathcal{N}(\mathcal{M}_{p,q} - \overline{y})|_{\mathcal{N}_{\overline{z}}})$, then by directly calculation, we know $g \pitchfork_{x'} \mathcal{DN}_{z'}$ in $\mathbb{S}^n$ is equivalent to $\Phi_{z'} \pitchfork_{(x', j_{z'}(\zeta))} \Delta$ in $\mathbb{S}^n \times \mathbb{S}^n$, where Δ is the diagonal of $\mathbb{S}^n \times \mathbb{S}^n$.

Since W-SRCQ holds at $\overline{z} \in \widetilde{\mathcal{M}}_{p,q}$, we have $g \pitchfork_{\overline{x}} \mathcal{DN}_{\overline{z}}$ in $\mathbb{S}^n$. By the above discussion, we know it is equivalent to that $\Phi_{\overline{z}} \pitchfork_{\left(\overline{x}, (g(\overline{x}), 0)\right)} \Delta$ in $\mathcal{M}_{p,q} \times \mathcal{M}_{p,q}$. Noting that

$$j_{z'}((G(\overline{z}), 0)) = E(\psi_{z'}(G(\overline{z}), 0)) - y' = E((G(z'), 0)) - y' = G(z') - y' = g(x'),$$

by Theorem B.2, we know that $\Phi_{z'} \pitchfork_{(x', (G(\overline{z}), 0))} \Delta$ in $\mathbb{S}^n \times \mathbb{S}^n$ holds for all $z' \in \widetilde{\mathcal{U}}$. Therefore, still by the above discussion, we know $g \pitchfork_{x'} \mathcal{DN}_{z'}$ in $\mathbb{S}^n$ for all $z' \in \widetilde{\mathcal{U}}$, which is exactly W-SRCQ holding at z'. □

Finally, we prove the genericity of W-SRCQ property. For any $p, q \in \mathbb{Z}_+$ and a smooth map $g : \mathbb{X} \to \mathbb{S}^n$, taking $b \in \mathbb{S}^n$, we denote that $g_b(x) = g(x) + b$, $G_b(z) = g(x) + b + y$ for $z = (x, y) \in \mathbb{X} \times \mathbb{S}^n$.

Theorem 5 (restated). *For generic $b \in \mathbb{S}^n$, W-SRCQ holds at all $\overline{z} = (\overline{x}, \overline{y})$ such that $\mathbb{S}^n_+ \ni g_b(\overline{x}) \perp \overline{y} \in \mathbb{S}^n_-$.*

Proof As before, we define notions parametrized by $b \in \mathbb{S}^n$: $G(x, b) := g_b(x) = g(x) + b$. Noting that G is a submersion, we know $G \pitchfork \mathcal{M}_{p,0}$ in $\mathbb{S}^n$ for any $p \in \mathbb{Z}_+$. Therefore, for each $p \in \{1, 2, ..., n\}$, we have a generic set $B_p \subseteq \mathbb{S}^n$ such that $g_b \pitchfork \mathcal{M}_{p,0}$ for any $b \in B_p$. Taking $B = \cap_{p=1}^n B_p$, we get a generic set $B \subseteq \mathbb{S}^n$ such that for every $b \in B$, we have $g_b \pitchfork \mathcal{M}_{p,0}$ in $\mathbb{S}^n$ for each $p = 1, ...n$.

Now suppose $g_b(\overline{x}) \in \mathcal{M}_{p,0}$, $\overline{y} \in \mathcal{M}_{0,q}$ and $g_b(\overline{x}) \perp \overline{y}$. Checking the definition of $\mathcal{N}_{\overline{z}}$, we know $\mathcal{M}_{p,0} \subseteq \mathcal{N}_{\overline{z}}$ for $\overline{z} = (\overline{x}, \overline{y})$. Hence $g_b \pitchfork_{\overline{x}} \mathcal{DN}_{\overline{z}}$ for any $\overline{z} = (\overline{x}, \overline{y}) \in \widetilde{\mathcal{M}}_{p,q}$ such that $\mathcal{M}_{p,0} \ni g_b(\overline{x}) \perp \overline{y} \in \mathcal{M}_{0,q}$, which exactly means W-SRCQ holds at $\overline{z} = (\overline{x}, \overline{y})$. □

C Proofs of Convergence Results

The following is the proof of Proposition 7.

Proposition 7 (restated). *For an arbitrary $z \in \mathbb{X} \times \mathbb{S}^n$ with $G(z)$ having an IED $(\alpha, \beta, \gamma, p, q, P, \lambda)$ and $v = (v_x, v_y) \in \mathbb{X} \times \mathbb{S}^n$, the directional derivative of φ at z along v can be expressed as*

$$\varphi'(z, v) = \left\langle dF_z^*(F(z)), \begin{bmatrix} v_x \\ H_1 \end{bmatrix} \right\rangle + \left\langle \nabla g(x)^*(F_1(z)), \Pi_{\mathbb{S}^n_-}(H_2) \right\rangle + \left\langle \nabla g(x)^*(F_1(z)) + F_2(z), \Pi_{\mathbb{S}^n_+}(H_2) \right\rangle, \tag{100}$$

where $F_1(z) = \nabla_x L(x, y) = \nabla f(x) + \langle y, \nabla g(x) \rangle$, $F_2(z) = -g(x) + \Pi_{\mathbb{S}^n_+}(G(z)) = -g(x) + \Pi_{\mathbb{S}^n_+}(g(x) + y)$ and $\begin{bmatrix} v_x \\ H_1 \end{bmatrix}$, H_2 are defined in Lemma 4.

Proof By the chain rule of B-differentiable functions, we have

$$\begin{aligned} \varphi'(z, v) &= F(z)^\top F'(z, v) \\ &= \left\langle F(z), \begin{bmatrix} \left(\nabla^2_{xx} L(z) - \nabla g(x) \nabla g(x)^*\right)(v_x) + \nabla g(x)(H) \\ -\nabla g(x)^*(v_x) + \Pi'_{\mathbb{S}^n_+}(G(z), H) \end{bmatrix} \right\rangle. \end{aligned}$$

Staring at equation (5), we have

$$\begin{aligned}\Pi'_{\mathbb{S}^n_+}(G(z),H) &= P\begin{bmatrix}\widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \Xi_{\alpha\gamma}\circ\widetilde{H}_{\alpha\gamma}\\ \widetilde{H}^\top_{\alpha\beta} & 0 & 0\\ \Xi^\top_{\alpha\gamma}\circ\widetilde{H}^\top_{\alpha\gamma} & 0 & 0\end{bmatrix}P^\top + P\begin{bmatrix}0 & 0 & 0\\ 0 & \Pi_{\mathcal{S}^{|\beta|}_+}(\widetilde{H}_{\beta\beta}) & 0\\ 0 & 0 & 0\end{bmatrix}P^\top\\ &= \Pi'_{\mathbb{S}^n_+}(G(z),H_1)+\Pi'_{\mathbb{S}^n_+}(G(z),H_2),\end{aligned}$$

where $H_1 = P\begin{bmatrix}\widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \widetilde{H}_{\alpha\gamma}\\ \widetilde{H}_{\beta\alpha} & 0 & \widetilde{H}_{\beta\gamma}\\ \widetilde{H}_{\gamma\alpha} & \widetilde{H}_{\gamma\beta} & \widetilde{H}_{\gamma\gamma}\end{bmatrix}P^\top \in \mathcal{T}_{G(z)}\mathcal{M}$ and $H_2 = P\begin{bmatrix}0 & 0 & 0\\ 0 & \widetilde{H}_{\beta\beta} & 0\\ 0 & 0 & 0\end{bmatrix}P^\top \in (\mathcal{T}_{G(z)}\mathcal{M})^\perp$ and $H = H_1 + H_2$.
Plugging this in, we get

$$\varphi'(z,v) = \left\langle F(z), \begin{bmatrix}\left(\nabla^2_{xx}L(z) - \nabla g(x)\nabla g(x)^*\right)(v_x) + \nabla g(x)(H)\\ -\nabla g(x)^*(v_x) + \Pi'_{\mathbb{S}^n_+}(G(z),H)\end{bmatrix}\right\rangle \tag{101}$$

$$= \left\langle F(z), \begin{bmatrix}\left(\nabla^2_{xx}L(z) - \nabla g(x)\nabla g(x)^*\right)(v_x) + \nabla g(x)(H_1)\\ -\nabla g(x)^*(v_x) + \Pi'_{\mathbb{S}^n_+}(G(z),H_1)\end{bmatrix}\right\rangle + \left\langle F(z), \begin{bmatrix}\nabla g(x)(H_2)\\ \Pi'_{\mathbb{S}^n_+}(G(z),H_2)\end{bmatrix}\right\rangle \tag{102}$$

$$= \left\langle F(z), \begin{bmatrix}\nabla^2_{xx}L - \nabla g(x)\nabla g(x)^* & \nabla g(x)\\ -\nabla g(x)^* & \xi_{G(z)}\end{bmatrix}\begin{bmatrix}v_x\\ H_1\end{bmatrix}\right\rangle + \left\langle F(z), \begin{bmatrix}\nabla g(x)(H_2)\\ \Pi_{\mathbb{S}^n_+}(H_2)\end{bmatrix}\right\rangle \tag{103}$$

$$= \left\langle F(z), dF_z\left(\begin{bmatrix}v_x\\ H_1\end{bmatrix}\right)\right\rangle + \left\langle F(z), \begin{bmatrix}\nabla g(x)(H_2)\\ \Pi'_{\mathbb{S}^n_+}(G(z),H_2)\end{bmatrix}\right\rangle \tag{104}$$

$$= \left\langle dF^*_z(F(z)), \begin{bmatrix}v_x\\ H_1\end{bmatrix}\right\rangle + \left\langle \nabla g(x)^*(F_1(z)), H_2\right\rangle + \left\langle F_2(z), \Pi_{\mathbb{S}^n_+}(H_2)\right\rangle \tag{105}$$

$$= \left\langle dF^*_z(F(z)), \begin{bmatrix}v_x\\ H_1\end{bmatrix}\right\rangle + \left\langle \nabla g(x)^*(F_1(z)), \Pi_{\mathbb{S}^n_-}(H_2)\right\rangle + \left\langle \nabla g(x)^*(F_1(z)) + F_2(z), \Pi_{\mathbb{S}^n_+}(H_2)\right\rangle, \tag{106}$$

where $F_1(z) = \nabla_x L(x,y) = \nabla f(x) + \langle y, \nabla g(x)\rangle$ and $F_2(z) = -g(x) + \Pi_{\mathbb{S}^n_+}(G(z)) = -g(x) + \Pi_{\mathbb{S}^n_+}(g(x)+y)$. □

Finally, we prove Lemma 5.

Lemma 5 (restated). *Given some $z = (x,y) \in \mathbb{X}\times\mathbb{S}^n$. If $W_1(z) \neq 0$, then it implies that $\|\nabla g(x)W_1(z)\| \neq 0$ and we have*

$$\operatorname*{argmin}_{t\geq 0}\ \varphi\left(z + t\begin{bmatrix}0\\ W_1(z)\end{bmatrix}\right) = \frac{\|W_1(z)\|^2}{\|\nabla g(x)W_1(z)\|^2}$$

and

$$\varphi(z) - \min_{t\geq 0}\varphi\left(z + t\begin{bmatrix}0\\ W_1(z)\end{bmatrix}\right) = \frac{1}{2}\frac{\|W_1(z)\|^4}{\|\nabla g(x)W_1(z)\|^2}.$$

Similarly, if $W_2(z) \neq 0$, we have

$$\operatorname*{argmin}_{t\geq 0}\ \varphi\left(z + t\begin{bmatrix}0\\ W_2(z)\end{bmatrix}\right) = \frac{\|W_2(z)\|^2}{\|W_2(z)\|^2 + \|\nabla g(x)W_2(z)\|^2}$$

and

$$\varphi(z) - \min_{t\geq 0}\varphi\left(z + t\begin{bmatrix}0\\ W_2(z)\end{bmatrix}\right) = \frac{1}{2}\frac{\|W_2(z)\|^4}{\|W_2(z)\|^2 + \|\nabla g(x)W_2(z)\|^2}.$$

Moreover, suppose Ω is a sufficiently large compact domain containing the whole area the algorithm is running, denoting the Lipschitz constant of g on Ω by $L_{g,\Omega}$ (due to the smoothness of g on $\mathbb{X}$), we have

$$\varphi(z) - \min_{t\geq 0}\varphi\left(z + t\begin{bmatrix}0\\ W_1(z)\end{bmatrix}\right) \geq \frac{1}{2L^2_{g,\Omega}}\|W_1(z)\|^2 \text{ or } \varphi(z) - \min_{t\geq 0}\varphi\left(z + t\begin{bmatrix}0\\ W_2(z)\end{bmatrix}\right) \geq \frac{1}{2(L^2_{g,\Omega}+1)}\|W_2(z)\|^2,$$

if $W_1(z) \neq 0$ or $W_2(z) \neq 0$, respectively.

Proof Note that $W_1(z)$ and $W_2(z)$ are both in $\mathcal{N}_{G(z)}\mathcal{M}_{p,q}$. So we can separate the directional derivative as what we did in the calculation of $\varphi'(z,v)$.

$$\begin{aligned}
2\varphi\left(z+t\begin{bmatrix}0\\W_1(z)\end{bmatrix}\right) &= \left\|F\left(z+t\begin{bmatrix}0\\W_1(z)\end{bmatrix}\right)\right\|^2\\
&= \left\|\begin{bmatrix}\nabla f(x)+\nabla g(x)(y+tW_1(z))\\ \Pi'_{\mathbb{S}^n_+}(G(z),H+tW_1(z))\end{bmatrix}\right\|^2\\
&= \left\|F(z)+t\begin{bmatrix}\nabla g(x)W_1(z)\\ \Pi'_{\mathbb{S}^n_+}(G(z),W_1(z))\end{bmatrix}\right\|^2\\
&= \|F(z)\|^2+2t\left\langle F(z),\begin{bmatrix}\nabla g(x)W_1(z)\\ \Pi'_{\mathbb{S}^n_+}(G(z),W_1(z))\end{bmatrix}\right\rangle + t^2\left\|\begin{bmatrix}\nabla g(x)W_1(z)\\ \Pi'_{\mathbb{S}^n_+}(G(z),W_1(z))\end{bmatrix}\right\|^2.
\end{aligned}$$

Since $W_1(z)\in\mathbb{S}^n_-$, we know $\Pi'_{\mathbb{S}^n_+}(G(z),W_1(z))=0$. Then direct computation shows

$$\left\langle F(z),\begin{bmatrix}\nabla g(x)W_1(z)\\ \Pi'_{\mathbb{S}^n_+}(G(z),W_1(z))\end{bmatrix}\right\rangle = -\|W_1(z)\|^2,\ \left\|\begin{bmatrix}\nabla g(x)W_1(z)\\ \Pi'_{\mathbb{S}^n_+}(G(z),W_1(z))\end{bmatrix}\right\|^2 = \|\nabla g(x)W_1(z)\|^2.$$

Thus the conclusion will follow from basic properties of quadratic functions if we know $\|\nabla g(x)W_1(z)\|^2\neq 0$.

Now we prove $\|\nabla g(x)W_1(z)\|\neq 0$ this by contradiction. Otherwise suppose $\|\nabla g(x)W_1(z)\|^2=0$. Then by definition of W_1, we have $\nabla g(x)\Pi_{\mathbb{S}^n_-}(-\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*(F_1(z))))=0$. Then after taking inner product with $F_1(z)$, applying orthogonal projection $-\pi^2_{z,\mathbb{S}^n}$ and separating into $\Pi_{\mathbb{S}^n_+}$-part and $\Pi_{\mathbb{S}^n_-}$-part, we get

$$\begin{aligned}
0 &= \left\langle F_1(z),\nabla g(x)\Pi_{\mathbb{S}^n_-}(-\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*(F_1(z))))\right\rangle\\
&= \left\langle \nabla g(x)^*F_1(z),\Pi_{\mathbb{S}^n_-}(-\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*(F_1(z))))\right\rangle\\
&= \left\langle -(\pi^1_{z,\mathbb{S}^n}+\pi^2_{z,\mathbb{S}^n})(\nabla g(x)^*F_1(z)),\Pi_{\mathbb{S}^n_-}(-\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*(F_1(z))))\right\rangle \text{(since } \pi^1_{z,\mathbb{S}^n}+\pi^2_{z,\mathbb{S}^n}=id\text{)}\\
&= \left\langle -\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*F_1(z)),\Pi_{\mathbb{S}^n_-}(-\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*(F_1(z))))\right\rangle \text{(since } R(\pi^1_{z,\mathbb{S}^n})=\mathcal{T}_{G(z)}\mathcal{M}\perp R(\pi^2_{z,\mathbb{S}^n})=(\mathcal{T}_{G(z)}\mathcal{M})^\perp\text{)}\\
&= \left\langle (\Pi_{\mathbb{S}^n_+}+\Pi_{\mathbb{S}^n_-})(-\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*F_1(z))),\Pi_{\mathbb{S}^n_-}(-\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*(F_1(z))))\right\rangle\\
&= \left\langle \Pi_{\mathbb{S}^n_-}(-\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*F_1(z))),\Pi_{\mathbb{S}^n_-}(-\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*(F_1(z))))\right\rangle \text{(since } \Pi_{\mathbb{S}^n_-}(A)\perp\Pi_{\mathbb{S}^n_+}(A) \text{ for any symmetric } A\text{)}\\
&= \|W_1(z)\|^2,
\end{aligned}\tag{107}$$

which contradicts to our assumption $W_1(z)\neq 0$.

Similar and easier argument can be applied to $W_2(z)$ since we have $\|W_2(z)\|^2+\|\nabla g(x)W_2(z)\|^2\neq 0$ automatically. □

Next, we prove the global convergence result of Algorithm 1, namely, Theorem 9. Recall that we define the lower modulus of eigenvalues at z in (70) by

$$\delta(z) := \min\{|\lambda(G(z))| : \lambda(G(z))\neq 0\}.$$

Theorem 9 (restated). *Taking tolerance to be $\epsilon=0$ and starting from an arbitrary initial point z^0, Algorithm 1 will stop in finite steps or we get an infinite sequence. For the finite steps case, Algorithm 1 will stop at a D-stationary point of $\varphi(z)=\frac{1}{2}\|F(z)\|^2$. For the infinite sequence case, any accumulation point z^∞ satisfying $\delta(z^\infty)>\delta$ of the sequence is a D-stationary point of $\varphi(z)$, where $\delta(z^\infty)$ is defined in (70) and δ is the parameter in Algorithm 1.*

In the following lemmas, we always assume that $\widetilde{\Omega}\subseteq\widetilde{\mathcal{M}}$ is some compact domain in some piece of stratum $\widetilde{\mathcal{M}}$ and $\Omega := \{(z,v)\mid z\in\widetilde{\Omega}, v\in\mathcal{T}_z\widetilde{\mathcal{M}}, \|v\|\leq l\}$ is the corresponding disk tangent bundle over $\widetilde{\Omega}$, where l is a fixed positive scalar. Hence Ω is compact. Moreover, because of incompleteness of $\widetilde{\mathcal{M}}$, we need to take $\widetilde{\Omega}$ and Ω small enough so that $\exp(\Omega)\subseteq\widetilde{\mathcal{M}}$ if necessary.

The following lemma shows that the directional derivative of φ along either of the stratum-LM direction and the two normal directions has the same order of the square of the norm of the direction.

Lemma C.1. *If $\mu\in[\overline{\mu},\mu_{\max}]$ is bounded, then there exists constants $M>m>0$ such that for the `SLMN` step at any $z\in\widetilde{\Omega}$ and any corresponding $d\in\{W_1(z),W_2(z),v^{LM}(z)\}$, we all have*

$$m\|d\|^2 < -\varphi'(z,d) < M\|d\|^2.$$

Proof If $d = W_2(z)$, then $d \in (\mathcal{T}_z\widetilde{\mathcal{M}})^\perp$ and we have

$$\begin{aligned}\varphi'(z,d) &= \langle \nabla g(x)^*(F_1(z)), \Pi_{\mathbb{S}^n_-}(W_2(z))\rangle + \langle \nabla g(x)^*(F_1(z)) + F_2(z), \Pi_{\mathbb{S}^n_+}(W_2)\rangle \\ &= \langle \nabla g(x)^*(F_1(z)) + F_2(z), W_2(z)\rangle \\ &= \langle \nabla g(x)^*(F_1(z)) + F_2(z), \Pi_{\mathbb{S}^n_+}(-(\nabla g(x)^*(F_1(z)) + F_2(z)))\rangle \\ &= -\|\Pi_{\mathbb{S}^n_+}(-(\nabla g(x)^*(F_1(z)) + F_2(z)))\|^2 \\ &= -\|W_2(z)\|^2 = -\|d\|^2.\end{aligned}$$

Similarly, if $d = W_1(z)$, we still have $\varphi'(z,d) = -\|d\|^2$.

On the other hand, if $d = v^{\mathrm{LM}}(z)$, then $d \in \mathcal{T}_z\widetilde{\mathcal{M}}$ and we have

$$\begin{aligned}\varphi'(z,d) &= -F(z)^\top dF_z(\mu(z) + dF_z^* dF_z)^{-1} dF_z^* F(z) \\ &= -(v^{\mathrm{LM}}(z))^\top(\mu(z) + dF_z^* dF_z)v^{\mathrm{LM}}(z) \\ &\le -\overline{\mu}\|v^{\mathrm{LM}}(z)\|^2 = -\overline{\mu}\|d\|^2,\end{aligned}$$

and $\varphi'(z,d) \ge -M'\|v^{\mathrm{LM}}(z)\|^2 = -M'\|d\|^2$ for any $M' > \mu_{\max} + \max\limits_{z\in\widetilde{\Omega}} \|dF_z\|^2$. So $\overline{\mu}\|v^{\mathrm{LM}}(z)\|^2 < -\varphi'(z, v^{\mathrm{LM}}(z)) < M'\|v^{\mathrm{LM}}(z)\|^2$.

Now take $m = \min\{1, \overline{\mu}\}$ and $M = \max\{1, M'\}$. So for whatever choice of d we take, we always have that

$$m\|d\|^2 < -\varphi'(z,d) < M\|d\|^2.$$

This completes the proof. □

The following lemma comes from the smoothness of retraction $R_{\widetilde{\mathcal{M}}}$ and the exponential map, packaging three things:

- For any compact domain $\widetilde{\Omega} \subseteq \widetilde{\mathcal{M}}_{p,q}$, we have a uniform constant $l_{\widetilde{\Omega}}$ only depending on $\widetilde{\Omega}$ such that once a tangent vector is shorter than this constant, applying the retraction or the exponential map to it would not lead outside the stratum;
- The retraction we construct differs from the exponential map by only a second order term;
- Moreover, the coefficient of this second order term is also uniformly bounded in $\widetilde{\Omega}$.

Lemma C.2. *For an arbitrary compact domain $\widetilde{\Omega} \subseteq \widetilde{\mathcal{M}}$, there exists a positive scalar $l_{\widetilde{\Omega}} > 0$ such that for $\Omega := \{(z,v) \mid z \in \widetilde{\Omega}, v \in \mathcal{T}_z\widetilde{\mathcal{M}}, \|v\| \le l_{\widetilde{\Omega}}\} \subseteq \mathcal{T}\widetilde{\mathcal{M}}$, we have three following properties:*

(i) $R_{\widetilde{\mathcal{M}}}(\Omega) \subseteq \widetilde{\mathcal{M}}$, *where $R_{\widetilde{\mathcal{M}}}$ is the retraction defined in (66);*
(ii) $\exp(\Omega) \subseteq \widetilde{\mathcal{M}}$;
(iii) There exists a constant κ_z only depending on z such that for any $(z,v) \in \Omega$,

$$\|R_z(v) - (z+v)\| \le \kappa_z\|v\|^2.$$

(iv) Moreover, we can obtain a constant $\kappa_{\widetilde{\Omega}}$ only depending on $\widetilde{\Omega}$ such that

$$\|R_z(v) - (z+v)\| \le \kappa_{\widetilde{\Omega}}\|v\|^2.$$

Proof This is a direct result from the smoothness of $R_{\widetilde{\mathcal{M}}}$ and the exponential map. □

With the above two lemmas, we prove in the following proposition and corollary that at a point $z \in \widetilde{\mathcal{M}}$, each stratum-LM step in the `SLMN` step will lead to a decreasing at least $O(\varepsilon_z\|v^{\mathrm{LM}}\|^2)$.

Proposition C.1. *As before, we assume $\mu \in [\overline{\mu}, \mu_{\max}]$ and drop the indices for simplicity. Suppose $z \in \widetilde{\Omega} \subseteq \widetilde{\mathcal{M}}$ for some fixed compact $\widetilde{\Omega}$ in some fixed $\widetilde{\mathcal{M}}$ and we get $\Omega \subseteq \mathcal{T}\widetilde{\mathcal{M}}$ under the rule of Lemma C.2. Then there exists a constant ε_z depending on the current point z such that for $v = v^{LM}(z)$ and any $t \in [0, \varepsilon_z)$, we have*

$$\varphi(z) - \varphi(R_z(tv)) \ge -\eta t\varphi'(z,v).$$

Proof By the smoothness of f, g and $\Pi_{\mathbb{S}^n_+}$ on $\mathcal{M}$ corresponding to $G(z)$, φ is smooth on $\widetilde{\mathcal{M}}$ and $v^{\mathrm{LM}}(z) := -(dF_z^* dF_z + \mu(z))^{-1} dF_z^*(F(z))$ is also a smooth mapping from $\widetilde{\mathcal{M}}$ to $\mathbb{S}^n$. So its norm bound

$$V(\widetilde{\Omega}) := \{\|v^{\mathrm{LM}}(z)\| \mid z \in \widetilde{\Omega}\}$$

only depends on $\widetilde{\Omega}$. So for $t \in \left(0, \dfrac{l_{\widetilde{\Omega}}}{V(\widetilde{\Omega})}\right]$, we always have $(z, tv^{\mathrm{LM}}(z)) \in \Omega$. For such t, by the smoothness of φ, we have

$$|\varphi(\exp_z(tv)) - \varphi(z) - d\varphi_z(tv)| = |\varphi(\exp_z(tv)) - \varphi(z) - \varphi'(z, tv)| \le C_\Omega \|tv\|^2, \tag{108}$$

and

$$\|\exp_z(tv) - (z + tv)\| \le C'_\Omega \|tv\|^2,$$

where C_Ω, C'_Ω only depend on φ, Ω. By the Lipschitz property of φ, we have

$$\begin{aligned} |\varphi(R_z(tv)) - \varphi(\exp_z(tv))| &\le \|R_z(tv) - \exp_z(tv)\| \\ &\le \|R_z(tv) - (z + tv)\| + \|\exp_z(tv) - (z + tv)\| \\ &\le (\kappa_z + C'_\Omega)\|tv\|^2. \end{aligned} \tag{109}$$

Combining (108), (109), we get that

$$|\varphi(R_z(tv)) - \varphi(z) - d\varphi_z(tv)| = |\varphi(R_z(tv)) - \varphi(z) - \varphi'(z, tv)| \le (\kappa_z + C'_\Omega + C_\Omega)\|tv\|^2.$$

Using the estimate in Lemma C.1, we get

$$\begin{aligned} \varphi(z) - \varphi(R_z(tv)) &\ge -\varphi'(z, tv) + (\kappa_z + C'_\Omega + C_\Omega)\|tv\|^2 \\ &= -t\varphi'(z, v) + (\kappa_z + C'_\Omega + C_\Omega)\|tv\|^2 \\ &\ge -t(1 - \frac{\kappa_z + C'_\Omega + C_\Omega}{M} t)\varphi'(z, v). \end{aligned}$$

Therefore, by taking $\varepsilon_z = \min\left\{\dfrac{(1-\eta)M}{\kappa_z + C'_\Omega + C_\Omega}, \dfrac{l_{\widetilde{\Omega}}}{V(\widetilde{\Omega})}\right\}$, wo obtain the desired result. □

Proposition C.2. *With the assumption in Proposition C.1, the stepsize in the Armijo line search along the stratum-LM direction of the* `SLMN` *step satisfies that $\rho^j > \rho\varepsilon_z$. Moreover, we have*

$$\varphi(z) - \varphi(R_z(\rho^j v)) \ge \eta\rho\varepsilon_z\overline{\mu}\|v\|^2.$$

Proof It is directly given by the definition of j, once $\rho^j < \varepsilon_z$, the Armijo condition will be satisfied and the iteration will stop. So the step before stopping has to satisfy $\rho^{j-1} > \varepsilon_z$, i.e. $\rho^j > \rho\varepsilon_z$. Then, by Proposition C.1, Lemma C.1, and the above estimate of the stepsize, we obtain the desired result. □

The following lemma tells that on the stratum where the accumulating point z^∞ stays, ε_z is lower-bounded by $\frac{1}{2}\varepsilon_{z^\infty}$. So part of the decrement has been controlled by data of z^∞.

Lemma C.3. *Suppose a sequence $\{z^k\} \subseteq \widetilde{\mathcal{M}}$ convergences to $z^\infty \in \widetilde{\mathcal{M}}$. Taking a compact neighborhood $\widetilde{\Omega}$ of z^∞ in $\widetilde{\mathcal{M}}$, we can define a compact Ω following the rule of Lemma C.2. Then when k is large enough, we have $\varepsilon_{z^k} > \dfrac{1}{2}\varepsilon_{z^\infty}$.*

Proof $\{z^k\}$ and z are all on the same $\widetilde{\mathcal{M}}$. So we have $|\alpha(G(z^k))| = |\alpha(G(z^\infty))|$, $|\beta(G(z^k))| = |\beta(G(z^\infty))|$ and $|\gamma(G(z^k))| = |\gamma(G(z^\infty))|$. Therefore, positive and negative eigenvalues of $G(z^k)$ can only convergence to the positive and negative eigenvalues of $G(z^\infty)$ respectively. This implies that $\delta_{z^k}^{-1} < 2\delta_{z^\infty}^{-1} + \dfrac{C_{g,\Omega} + C'_\Omega + C_\Omega}{2(1 + L_{g,\Omega}^2)}$ when k is large enough. By the definition of κ_z, we have $\kappa_{z^k} < 2\kappa_{z^\infty} + C'_\Omega + C_\Omega$, which in consequence implies $\varepsilon_{z^k} > \dfrac{1}{2}\varepsilon_{z^\infty}$. □

The following lemma shows that on the stratum where the accumulating point z^∞ stays, the decrement of the normal step (calculated in Lemma 5) in the `SLMN` step are all controlled by data of z^∞.

Lemma C.4. *If restricted on a fixed $\widetilde{\mathcal{M}}$,*

$$W_1(z) := \Pi_{\mathbb{S}^n_-}(-\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*(F_1(z)))),\ W_2(z) := \Pi_{\mathbb{S}^n_+}\big(-\pi^2_{z,\mathbb{S}^n}(\nabla g(x)^*(F_1(z)) + F_2(z))\big)$$

are both smooth mappings from $\widetilde{\mathcal{M}}$ into $\mathbb{S}^n$.

Proof This relies on the the smoothness of $\pi^2_{z,\mathbb{S}^n}$ with respect to z, which has been proven in Appendix A. □

Now we can give the proof of the global convergence.

Proof (Global convergence of Theorem 9) If the sequence terminates in finite steps, then the last point will be a D-stationary point, which is a trivial case.

Suppose $\{z^k\}_{k\in\theta} \subseteq \mathbb{Z}_+$ is a convergent subsequence of $\{z^k\}_{k\in\mathbb{Z}_+}$ and $\lim\limits_{k(\in\theta)\to\infty} z^k = z^\infty$ and we denote the manifold corresponding to z^∞ just by $\widetilde{\mathcal{M}}$. Take a compact neighborhood $\widetilde{\Omega} \subseteq \widetilde{\mathcal{M}}$ of z^∞ and a compact set $\Omega \subseteq \mathcal{T}\widetilde{\mathcal{M}}$ satisfying the conditions in Lemma C.2. Then there exists some $K > 0$ such that when $k(\in\theta) > K$, $\|z^k - z^\infty\|$ is small enough and by Lemma 6 $\|\widehat{z^k} - z^\infty\|$ is also small enough. Hence we have $\left\{\widehat{z^k}\right\}_{k(\in\theta)>K} \subseteq \widetilde{\Omega}$.

Since $\{\varphi(z^k)\}_{k\in\theta}$ is decreasing, we have

$$\sum_{k(\in\theta)>K} \varphi(z^k) - \varphi(z^{k+1}) \le \varphi(z^0) < +\infty. \tag{110}$$

If $\lim\limits_{k\in\to\infty} \mu_k = 0$, then z^∞ is a KKT solution point, which is exactly what we want. So we only need to consider the case that we get a decreasing and lower-bounded sequence $\mu_{\max} > \mu_0 > \mu_1 > ... > \overline{\mu} > 0$. Noting that F is Lipschitz over a compact domain in $\mathbb{X} \times \mathbb{S}^n$ containing the δ-neighborhood of the whole subsequence $\left\{z^k\right\}_{k\in\theta}$, we may take $\mu_{\max}$ large enough so that every point we may encounter will satisfy $\mu < \mu_{\max}$.

By the definition of z^{k+1}, if $\varphi(\texttt{SLMN}(\widehat{z^k})) \le \varphi(z^k)$, $z^{k+1} = \texttt{SLMN}(\widehat{z^k})$. Otherwise, $z^{k+1} = \texttt{SLMN}(z^k)$, in which case $\varphi(z^{k+1}) \le \varphi(z^k) < \varphi(\texttt{SLMN}(\widehat{z^k}))$. In either case, we both have

$$\begin{aligned}\varphi(z^k) - \varphi(z^{k+1}) &\ge \varphi(z^k) - \varphi(\texttt{SLMN}(\widehat{z^k}))\\ &\ge \varphi(\widehat{z^k}) - \varphi(\texttt{SLMN}(\widehat{z^k})) - |\varphi(z^k) - \varphi(\widehat{z^k})|\end{aligned}$$

And by the rule of SLMN step, we have

$$\varphi(\widehat{z^k}) - \varphi(\texttt{SLMN}(\widehat{z^k})) \ge \max\left\{\varphi(\widehat{z^k}) - \varphi(R_{\widehat{z^k}}\left(\widehat{v^k}\right)), \varphi(\widehat{z^k}) - \min_{t\ge0}\varphi(\widehat{z^k} + t\widehat{W_1^k}), \varphi(\widehat{z^k}) - \min_{t\ge0}\varphi(\widehat{z^k} + t\widehat{W_2^k})\right\},$$

where the $\widehat{\ }$ of $\widehat{v^k}, \widehat{W_1^k}, \widehat{W_2^k}$ means they evaluate at $\widehat{z^k}$. Since $\varphi(\widehat{z^k}) \ge \varphi(R_{\widehat{z^k}}\left(\widehat{v^k}\right)) \ge \varphi(z^{k+1})$, we know $\mu(\widehat{z^k}) > \overline{\mu}$. So Proposition C.1 and Proposition C.2 can be applied and we get

$$\varphi(\widehat{z^k}) - \varphi(R_{\widehat{z^k}}\left(\widehat{v^k}\right)) \ge \eta\rho\varepsilon_{\widehat{z^k}}\overline{\mu}\|\widehat{v^k}\|^2.$$

And Lemma 5 tells that

$$\varphi(\widehat{z^k}) - \min_{t\ge0}\varphi(\widehat{z^k} + t\widehat{W_1^k}) \ge \frac{1}{2L_{g,\Omega}^2}\|\widehat{W_1^k}\|^2 \quad\text{and}\quad \varphi(\widehat{z^k}) - \min_{t\ge0}\varphi(\widehat{z^k} + t\widehat{W_2^k}) \ge \frac{1}{2(L_{g,\Omega}^2+1)}\|\widehat{W_2^k}\|^2.$$

Therefore, we get

$$\varphi(\widehat{z^k}) - \varphi(\texttt{SLMN}(\widehat{z^k})) \ge \max\{\eta\rho\varepsilon_{\widehat{z^k}}\overline{\mu}\|\widehat{v^k}\|^2, \frac{1}{2L_{g,\Omega}^2}\|\widehat{W_1^k}\|^2, \frac{1}{2(L_{g,\Omega}^2+1)}\|\widehat{W_2^k}\|^2\}.$$

By the definition of $\widehat{z^k}$ and $\delta(z^\infty) > \delta$, we know $\widehat{z^k}$ and z^∞ are eventually on the same $\widetilde{\mathcal{M}}$ when $\|z^k - z^\infty\| \to 0$. Then by Lemma C.4, we have

$$\varphi(\widehat{z^k}) - \varphi(\texttt{SLMN}(\widehat{z^k})) > \max\left\{\frac{1}{4}\eta\rho\varepsilon_{z^\infty}\overline{\mu}\|v^\infty\|^2, \frac{1}{4L_{g,\Omega}^2}\|W_1^\infty\|^2, \frac{1}{4(L_{g,\Omega}^2+1)}\|W_2^\infty\|^2\right\} > \overline{\varepsilon}s(z^\infty,\varphi)^2,$$

where $\overline{\varepsilon} := \min\left\{\frac{1}{4}\eta\rho\varepsilon_{z^\infty}\overline{\mu}, \frac{1}{4L_{g,\Omega}^2}, \frac{1}{4(L_{g,\Omega}^2+1)}\right\}$ and $W_1^\infty = W_1(z^\infty), W_2^\infty = W_2(z^\infty)$.

Since z^k converges to z^∞, Lemma 6 tells that $\widehat{z^k}$ also converges to z^∞. Note that $\overline{\varepsilon}s(z^\infty,\varphi)^2$ is a constant. So when k is large enough,

$$|\varphi(z^k) - \varphi(\widehat{z^k})| \le |\varphi(z^k) - \varphi(z^\infty)| + |\varphi(\widehat{z^k}) - \varphi(z^\infty)| \le \frac{1}{2}\overline{\varepsilon}s(z^\infty,\varphi)^2.$$

From (110) and the monotonicity of $\varphi(z^k)$, we know that

$$\sum_{k(\in\theta)>K} \frac{1}{2}\overline{\varepsilon} s(z^\infty,\varphi)^2 \le \sum_{k(\in\theta)>K} \varphi(z^k) - \varphi(z^{k+1}) \le \varphi(z^0) < +\infty,$$

which implies that $\frac{1}{2}\overline{\varepsilon} s(z^\infty,\varphi)^2 = 0$. Therefore, by the definition of $s(z^\infty,\varphi)$, z^∞ is a D-stationary point of φ. □

Next, we provide the proof of the results establishing the local quadratic convergence rate of Algorithm 1.

Lemma 7 (restated). *Let $\overline{z} = (\overline{x},\overline{y})$ be a KKT pair of (1) with IED $(\alpha,\beta,\gamma,p,q,\overline{P},\overline{\lambda})$ of $G(\overline{z})$. If the W-SOC and the SRCQ holds at $\overline{z}$, then 0 is the only zero of $F(z + \begin{bmatrix}0\\ \cdot\end{bmatrix})$ on $\mathcal{N}_{G(\overline{z})}\mathcal{M}_{p,q}$.*

Proof By (72), we know 0 is not the only zero if and only if there exists $0 \neq \Delta \in_{G(\overline{z})} \mathcal{M}_{p,q}$ such that $J_{\overline{z}}(\Delta) = 0$. For following special choice of element in $\mathbb{S}^n$

$$Y = P \begin{bmatrix} 0 & 0 & 0 \\ 0 & -I_{|\beta|} & 0 \\ 0 & 0 & 0 \end{bmatrix} P^\top,$$

by the SRCQ, there exist $x \in \mathbb{X}$ and $Y_0 \in \mathcal{T}_{\mathbb{S}^n_+}(\overline{x}) \cap \overline{y}^\perp$ such that

$$\nabla g(\overline{x})^* x + Y_0 = Y.$$

Here, we have the following characterization

$$\mathcal{T}_{\mathbb{S}^n_+}\big(g(\overline{x})\big) \cap \overline{y}^\perp = \left\{ \overline{P} B \overline{P}^\top \in \mathbb{S}^n \mid B_{\beta\beta} \succeq 0,\ B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0 \right\}. \tag{111}$$

Noticing that $0 \neq \Delta \in \mathcal{N}_{G(\overline{z})}\mathcal{M}_{p,q}$, we have

$$\begin{aligned} 0 <& \langle Y, \Delta\rangle && (112)\\ =& \langle \nabla g(\overline{x})^* x, \Delta\rangle + \langle Y_0, \Delta\rangle && (113)\\ =& \langle x, \nabla g(\overline{x})\Delta\rangle + \langle Y_0, \Delta\rangle \le 0, && (114) \end{aligned}$$

where the first inequality follows from $\Pi_{\mathbb{S}^n_+}(\Delta) = 0$ and the last inequality further follows from $\nabla g(\overline{x})\Delta = 0$ and (111). The contradiction leads to the desired isolatedness of the solution. □

Lemma 8 (restated). *For any z, J_z is positively homogeneous. If the W-SOC and the SRCQ holds at a KKT pair $\overline{z}$, there exist constants $0 < c_1 < c_2$ such that*

$$c_1\|\Delta\| \le \|J_{\overline{z}}(\Delta)\| \le c_2\|\Delta\| \quad \forall\, \Delta \in \mathcal{N}_{G(\overline{z})}\mathcal{M}_{p,q}, \tag{115}$$

where $\mathcal{M}_{p,q}$ is given by $\overline{z}$.

Proof The positive homogeneousness of J_z is obvious and the existence of c_2 follows directly from the Lipschitz property of $J_{\overline{z}}$. If there does not exist such a c_1, then since $J_{\overline{z}}$ is positively homogeneous, there exists a sequence $\Delta^\nu \in \mathcal{N}_{G(\overline{z})}\mathcal{M}_{p,q}$ such that $\|\Delta^\nu\| = 1$ and $\|J_{\overline{z}}(\Delta^\nu)\| < \frac{1}{\nu}$. It follows that there exists $\Delta^\infty \in \mathcal{N}_{G(\overline{z})}\mathcal{M}_{p,q}$ such that $\|\Delta^\infty\| = 1$ and $J_{\overline{z}}(\Delta^\infty) = 0$. This, by noting (72) and again the positive homogeneousness, contradicts Lemma 7. □

Lemma 9 (restated). *Let $\overline{z}$ be a KKT pair of (1) with an IED $(\alpha,\beta,\gamma,p,q,\overline{P},\overline{\lambda})$ of $G(\overline{z})$. If the W-SOC and the W-SRCQ holds at $\overline{z}$, then for $z \in \widetilde{\mathcal{M}}_{p,q}$ sufficiently close to $\overline{z}$ and $v = v^{LM}(z)$, there exists $0 < m < M$ such that*

$$m\|v\|^2 \le \|\varphi'(z,v)\| \le M\|v\|^2.$$

Proof Since $z \in \widetilde{\mathcal{M}}_{p,q}$, a direct computation shows that

$$\begin{aligned} \varphi'(z,v) &= -F(z)^\top dF_z(\mu(z) + dF_z^* dF_z)^{-1} dF_z^* F(z) && (116)\\ &= -v^\top(\mu(z) + dF_z^* dF_z)v, && (117) \end{aligned}$$

where $\mu(z) = \|F(z)\|^2$. Notice that, the linear operator $dF_z^* dF_z + \mu(z)$ is uniformly bounded when $z \in \widetilde{\mathcal{M}}_{p,q}$ is sufficiently close to $\overline{z}$, due to the smoothness of F. Moreover, by Corollary 2, $dF_{\overline{z}}^* dF_{\overline{z}}$ is positive definite. Consequently, $dF_z^* dF_z$ is uniformly positive definite in a neighborhood of $\overline{z}$, and so is $dF_z^* dF_z + \mu(z)$. These observations imply the desired conclusion. □

Proposition 10 (restated). *Let $\overline{z} = (\overline{x}, \overline{y})$ be a KKT pair of (1) with an IED $(\alpha, \beta, \gamma, p, q, \overline{P}, \overline{\lambda})$ of $G(\overline{z})$. If the W-SOC and the SRCQ hold at $\overline{z}$, then there exists a constant $c_3 > 0$ such that for any $z \in \widetilde{\mathcal{M}}_{p,q}$ sufficiently close to $\overline{z}$,*

$$\varphi(z) \leq c_3\, \varphi\left(z + t \begin{bmatrix} 0 \\ \Delta \end{bmatrix}\right), \tag{118}$$

for any $\Delta \in \mathcal{N}_{G(z)}\mathcal{M}_{p,q}$ and $t \in \mathbb{R}$.

Proof If we have the following estimate for the coefficient multiplying $\varphi(z)$ on the right-hand side of (73),

$$\limsup_{z \to \overline{z},\ z \in \widetilde{\mathcal{M}}_{p,q}} \left[\sup_{\Delta \in \mathcal{N}_{G(z)}\mathcal{M}_{p,q}} \langle \frac{F(z)}{\|F(z)\|}, \frac{J_z(\Delta)}{\|J_z(\Delta)\|} \rangle \right] < 1, \tag{119}$$

then the desired result (74) follows directly.

We will prove (119) by contradiction. Recall that when $F(z) = 0$ or $J_z(\Delta) = 0$, $\langle \frac{F(z)}{\|F(z)\|}, \frac{J_z(\Delta)}{\|J_z(\Delta)\|} \rangle$ is defined to be zero. Then, noting the positive homogeneity of J_z by Lemma 8, the negation of the claim is equivalent to saying there exist $z^\nu \in \widetilde{\mathcal{M}}_{p,q}$, $z^\nu \to \overline{z}$, $\Delta^\nu \in \mathcal{N}_{G(z^\nu)}\mathcal{M}_{p,q}$ and $\|\Delta^\nu\| = 1$ that

$$\lim_{\nu \to \infty} \langle \frac{F(z^\nu)}{\|F(z^\nu)\|}, \frac{J_{z^\nu}(\Delta^\nu)}{\|J_{z^\nu}(\Delta^\nu)\|} \rangle = 1,$$

which is equivalent to

$$\lim_{\nu \to \infty} \frac{F(z^\nu)}{\|F(z^\nu)\|} - \frac{J_{z^\nu}(\Delta^\nu)}{\|J_{z^\nu}(\Delta^\nu)\|} = 0. \tag{120}$$

For the first part, by the smoothness of F on $\widetilde{\mathcal{M}}_{p,q}$, we have that as $\nu \to \infty$,

$$\frac{F(z^\nu)}{\|F(z^\nu)\|} = \frac{F(\overline{z}) + dF_{\overline{z}}(w^\nu) + o(\|z^\nu - \overline{z}\|)}{\|F(z^\nu)\|} = dF_{\overline{z}}(\frac{w^\nu}{\|w^\nu\|}) \frac{\|w^\nu\|}{\|F(z^\nu)\|} + \frac{o(\|z^\nu - \overline{z}\|)}{\|F(z^\nu)\|}, \tag{121}$$

where $w^\nu = \exp_{z^\nu}^{-1}(\overline{z})$ and we used the fact that $\|w^\nu\|$ has the same order with $\|z^\nu - \overline{z}\|$ implied by the smoothness of the exponential map. Since $z^\nu \in \widetilde{\mathcal{M}}_{p,q}$, the stratum-restricted local error bound (36) and the Lipschitz property of F on $\mathcal{M}_{p,q}$ imply that $\|z^\nu - \overline{z}\|/\|F(z^\nu)\|$ is bounded below away from zero and bounded above. Taking a subsequence if necessary, we can assume that

$$\frac{w^\nu}{\|w^\nu\|} \to v \in \mathcal{T}_{\overline{z}}\widetilde{\mathcal{M}}_{p,q} \quad \text{and} \quad \frac{\|w^\nu\|}{\|F(z^\nu)\|} \to c_4, \tag{122}$$

where $\|v\| = 1$ and $c_4 > 0$. Putting (122) into (121), we have

$$\lim_{\nu \to \infty} \frac{F(z^\nu)}{\|F(z^\nu)\|} = c_4 dF_{\overline{z}}(v). \tag{123}$$

For the second part, without loss of generality, we assume that $\Delta^\nu \to \Delta^\infty$. Since $\|\Delta^\nu\| = 1$, we know $0 \neq \Delta^\infty \in \mathcal{N}_{G(\overline{z})}\mathcal{M}_{p,q}$. Then

$$\lim_{\nu \to \infty} J_{z^\nu}(\Delta^\nu) = \lim_{\nu \to \infty} \begin{bmatrix} \nabla g(x^\nu)\Delta^\nu \\ \Pi_{\mathbb{S}^n_+}(\Delta^\nu) \end{bmatrix} = \begin{bmatrix} \nabla g(\overline{x})\Delta^\infty \\ \Pi_{\mathbb{S}^n_+}(\Delta^\infty) \end{bmatrix} = J_{\overline{z}}(\Delta^\infty). \tag{124}$$

Combining (120), (123) and (124), set $s :=$ that

$$dF_{\overline{z}}(v) = sJ_{\overline{z}}(\Delta^\infty), \tag{125}$$

i.e.,

$$\begin{cases} (\nabla^2_{xx} L^* - \nabla g(\overline{x})\nabla g(\overline{x})^*)v_x + \nabla g(\overline{x})H - s\nabla g(\overline{x})\Delta^\infty = 0, \\ -\nabla g(\overline{x})^* v_x + \xi_{G(\overline{z})} H - s\Pi_{\mathbb{S}^n_+}(\Delta^\infty) = 0, \end{cases} \tag{126}$$

where $v = (v_x, H) \in \mathcal{T}_{\overline{x}}\mathbb{X} \times \mathcal{T}_{G(\overline{z})}\mathcal{M}_{p,q} \cong \mathcal{T}_{\overline{z}}\widetilde{\mathcal{M}}_{p,q}$ and we have made use of the characterization of $dF_{\overline{z}}(v)$ given in (25). The second equation in (126) implies that

$$(-\nabla g(\overline{x})^* v_x + \xi_{G(\overline{z})} H)_{\alpha\gamma} = 0 \quad \text{and} \qquad v_x \in \mathrm{appl}(\overline{z}), \tag{127}$$

where the set $\mathrm{appl}(\overline{z})$ is defined in (14) and characterized in (27).

Then, we have

$$\begin{aligned}0 =&\langle v_x, (\nabla^2_{xx}L^* - \nabla g(\overline{x})\nabla g(\overline{x})^*)v_x + \nabla g(\overline{x})H - s\nabla g(\overline{x})\Delta^\infty\rangle\\ =&\langle v_x, \nabla^2_{xx}L^* v_x\rangle + \langle \nabla g(\overline{x})^* v_x, H - s\Delta^\infty - \nabla g(\overline{x})^* v_x\rangle\\ =&\langle v_x, \nabla^2_{xx}L^* v_x\rangle + \langle \xi_{G(\overline{z})}H - s\Pi_{\mathbb{S}^n_+}(\Delta^\infty), H - s\Delta^\infty - \xi_{G(\overline{z})}H + s\Pi_{\mathbb{S}^n_+}(\Delta^\infty)\rangle\\ =&\langle v_x, \nabla^2_{xx}L^* v_x\rangle + \langle \xi_{G(\overline{z})}H, H - \xi_{G(\overline{z})}H\rangle - \langle s\Pi_{\mathbb{S}^n_+}(\Delta^\infty), -s\Delta^\infty + s\Pi_{\mathbb{S}^n_+}(\Delta^\infty)\rangle\\ =&\langle v_x, \nabla^2_{xx}L^* v_x\rangle + \langle \xi_{G(\overline{z})}H, H - \xi_{G(\overline{z})}H\rangle, \end{aligned} \tag{128}$$

where the fourth equation follows from $H \in \mathcal{T}_{G(\overline{z})}\mathcal{M}_{p,q}$, $\Delta^\infty \in \mathcal{N}_{G(\overline{z})}\mathcal{M}_{p,q}$ and definition of $\xi_{G(z)}$ in (20) and the fifth equation follows from the Moreau-Yosida decomposition of K.

Let $\widetilde{H} = P^\top HP$, we have

$$\langle \xi_{G(\overline{z})}H, H - \xi_{G(\overline{z})}H\rangle \tag{129}$$

$$=\langle P\begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & U_{\alpha\gamma}\circ\widetilde{H}_{\alpha\gamma}\\ \widetilde{H}^\top_{\alpha\beta} & 0 & 0\\ \widetilde{H}^\top_{\alpha\gamma}\circ U^\top_{\alpha\gamma} & 0 & 0\end{bmatrix}P^\top, P\begin{bmatrix} 0 & 0 & (E_{\alpha\gamma}-U_{\alpha\gamma})\circ\widetilde{H}_{\alpha\gamma}\\ 0 & \widetilde{H}_{\beta\beta} & \widetilde{H}_{\beta\gamma}\\ \widetilde{H}^\top_{\alpha\gamma}\circ(E_{\alpha\gamma}-U_{\alpha\gamma}) & \widetilde{H}^\top_{\beta\gamma} & \widetilde{H}_{\gamma\gamma}\end{bmatrix}P^\top\rangle \tag{130}$$

$$=2\langle U_{\alpha\gamma}\circ\widetilde{H}_{\alpha\gamma}, (E_{\alpha\gamma}-U_{\alpha\gamma})\circ\widetilde{H}_{\alpha\gamma}\rangle \tag{131}$$

$$=2\sum_{i\in\alpha}\sum_{j\in\gamma}\frac{-y_iy_j}{(y_i-y_j)^2}\widetilde{H}^2_{ij} \tag{132}$$

$$=2\sum_{i\in\alpha}\sum_{j\in\gamma}\frac{-y_iy_j}{(y_i-y_j)^2}(\frac{y_i-y_j}{y_i})^2(P^\top(-\nabla g(\overline{x})^* v_x)P)^2_{ij} = -\sigma_{\overline{z}}(v_x). \tag{133}$$

Here, the fourth equation follows from (127). Combine (133), (128), (127) and the W-SOC, we know $v_x = 0$. Therefore, (126) reduced to

$$\begin{cases}\nabla g(\overline{x})H - s\nabla g(\overline{x})\Delta^\infty = 0,\\ \xi_{G(\overline{z})}H - s\Pi_{\mathbb{S}^n_+}(\Delta^\infty) = 0.\end{cases} \tag{134}$$

As $H \in \mathcal{T}_{G(\overline{z})}\mathcal{M}_{p,q}$ and $\Delta^\infty \in \mathcal{N}_{G(\overline{z})}\mathcal{M}_{p,q}$, we must have $\Delta^\infty \in -K$, $\widetilde{H}_{\alpha\alpha} = 0$ and $\widetilde{H}_{\alpha\beta} = 0$. Then, for the following special choice of element in $\mathbb{S}^n$

$$Y = P\begin{bmatrix}0 & 0 & 0\\ 0 & -I_{|\beta|} & 0\\ 0 & 0 & 0\end{bmatrix}P^\top,$$

by the SRCQ, there exist $x \in \mathbb{X}$ and $Y_0 \in \mathcal{T}^n_{\mathbb{S}^n_+}(x)\cap y^\perp$ such that

$$\nabla g(\overline{x})^* x + Y_0 = Y.$$

Noticing that $0 \neq \Delta^\infty \in \mathcal{N}_{G(\overline{z})}\mathcal{M}_{p,q}\cap(-K)$, we have

$$0 >\langle Y, H - s\Delta^\infty\rangle \tag{135}$$

$$=\langle \nabla g(\overline{x})^* x, H - s\Delta^\infty\rangle + \langle Y_0, H - s\Delta^\infty\rangle \tag{136}$$

$$=\langle x, \nabla g(\overline{x})(H - s\Delta^\infty)\rangle + \langle Y_0, H\rangle - s\langle Y_0, \Delta^\infty\rangle \geq 0, \tag{137}$$

where the first inequality follows from $\Delta^\infty \in -K$ and the definition of Y, while the last inequality further follows from (134) and (111). All in all, the contradiction shows that the statement of this proposition holds and the proof is then completed. □

Proposition 11 (restated). *Suppose that $\overline{z} = (\overline{x}, \overline{y})$ is a KKT pair of (1) with an IED $(\alpha, \beta, \gamma, p, q, \overline{P}, \overline{\lambda})$ of $G(\overline{z})$. Then, for $z \in \widetilde{\mathcal{M}}_{p,q}$ sufficiently close to $\overline{z}$*

$$\varphi(R_z(v)) = \varphi(z) + \frac{1}{2}\varphi'(z, v) + o(\|v\|^2), \tag{138}$$

where $v = v^{LM}(z)$.

Proof By Corollary 2, the W-SOC and the SRCQ imply that when $z \in \widetilde{\mathcal{M}}_{p,q}$ is sufficiently close to $\overline{z}$, the norm $\left\|((dF_z)^* dF_z)^{-1}\right\|$ is uniformly bounded, and thus $\left\|\left((dF_z)^* dF_z + \|F(z)\|^2 I\right)^{-1}\right\|$, is also uniformly bounded. As a result, $v = v^{\mathrm{LM}}(z) \to 0$ when $z \to \overline{z}$.

By the smoothness of F on $\widetilde{\mathcal{M}}_{p,q}$, we can do Taylor expansion of $\varphi \circ R_z$ at $z \in \widetilde{\mathcal{M}}_{p,q}$ sufficiently close to $\overline{z}$:

$$\varphi(R_z(v)) = \varphi(z) + (\varphi \circ R_z)'(0; v) + \frac{1}{2}\nabla^2(\varphi \circ R_z)(0; v, v) + O(\|v\|^3), \tag{139}$$

since φ is of class C^∞ on $\widetilde{\mathcal{M}}_{p,q}$ and the retraction $R_{\widetilde{\mathcal{M}}_{p,q}} : \mathcal{T}\widetilde{\mathcal{M}}_{p,q} \to \widetilde{\mathcal{M}}_{p,q}$ is smooth by Proposition 9.

Since $d(R_z)_0 = id$, we have $(\varphi \circ R_z)'(0; v) = \varphi'(z; v) = F(z)^\top dF_z(v)$. For the 2nd-order term, directly computing Riemannian Hessian [1, Definition 5.5.1] of composition of maps between Riemannian manifolds,

$$\nabla^2(\varphi \circ R_z)(0; v, v) = \frac{1}{2}\left\{\mathrm{Hess}_{\widetilde{\mathcal{M}}_{p,q}}\varphi(z)(v, v) + d\varphi_z\left(\mathrm{Hess}^{\widetilde{\mathcal{M}}_{p,q}}_{\mathcal{T}_z\widetilde{\mathcal{M}}_{p,q}} R_z(0)(v, v)\right)\right\} \tag{140}$$

where $\mathrm{Hess}_{\widetilde{\mathcal{M}}_{p,q}}\varphi = \nabla(d\varphi)$ is the Riemannian Hessian of φ on $\widetilde{\mathcal{M}}_{p,q}$ and $\mathrm{Hess}^{\widetilde{\mathcal{M}}_{p,q}}_{\mathcal{T}_z\widetilde{\mathcal{M}}_{p,q}} R_z = \nabla(dR_z)$ is the second fundamental form ([26, Definition 5.1.2]) of $R_z : \mathcal{T}_z\widetilde{\mathcal{M}}_{p,q} \to \widetilde{\mathcal{M}}_{p,q}$. For the first term in (140), we have

$$\begin{aligned}
\mathrm{Hess}_{\widetilde{\mathcal{M}}_{p,q}}\varphi(v, v) &= (dF_z(v))^\top dF_z(v) + F(z)^\top \mathrm{Hess}_{\widetilde{\mathcal{M}}_{p,q}} F(z)(v, v) && (141)\\
&= -F(z)^\top dF_z(v) - \mu(z)\|v\|^2 + O\left(\|F(z)\|\|v\|^2\right) && (142)\\
&= -\varphi'(z, v) + o\left(\|v\|^2\right), && (143)
\end{aligned}$$

where $\mu(z) = \|F(z)\|^2$ is the regularization term in the computation of $v = v^{\mathrm{LM}}(z)$. Here, the first equality comes from direct calculation on Riemannian manifold $\widetilde{\mathcal{M}}_{p,q}$, the second is implied by the definition of v and the smoothness of F on $\widetilde{\mathcal{M}}_{p,q}$, and the last results from $\lim_{k\to\infty}\|F(z)\| = 0$, which is our assumption. As for the second term in (140), by the smoothness shown in Proposition 9, $\mathrm{Hess}^{\widetilde{\mathcal{M}}_{p,q}}_{\mathcal{T}_z\widetilde{\mathcal{M}}_{p,q}} R_z(0)(v, v) = O\left(\|v\|^2\right)$. So

$$d\varphi_z\left(\mathrm{Hess}^{\widetilde{\mathcal{M}}_{p,q}}_{\mathcal{T}_z\widetilde{\mathcal{M}}_{p,q}} R_z(0)(v, v)\right) = O\left(\|F(z)\|\|v\|^2\right) = o\left(\|v\|^2\right). \tag{144}$$

Combining (139), (140), (143) and (144), we then obtain the desired result immediately. □

Proposition 12 (restated). *Let $\overline{z}$ be a KKT pair of (1) with an IED $(\alpha, \beta, \gamma, p, q, \overline{P}, \overline{\lambda})$ of $G(\overline{z})$. If the W-SOC and the W-SRCQ holds at $\overline{z}$, then for $z \in \widetilde{\mathcal{M}}_{p,q}$ sufficiently close to $\overline{z}$, we have*

$$\left\|R_z(v) - \overline{z}\right\| = O\big(\|z - \overline{z}\|^2\big),$$

where R_z is the retraction of $\widetilde{\mathcal{M}}_{p,q}$ at z given in (66) and $v = v^{LM}(z)$ is given by (65).

Proof By Lemma C.2, the retraction R_z admits a second order expansion. Thus, for all z sufficiently close to $\overline{z}$,

$$\left\|R_z(v) - \overline{z}\right\| = \left\|z - \overline{z} - \left((dF_z)^* dF_z + \mu(z)I\right)^{-1}(dF_z)^* F(z)\right\| + O(\|v\|^2),$$

where $\mu(z) = \|F(z)\|^2$ is the regularization parameter used in the definition of the Levenberg–Marquardt step $v = v_{\mathrm{LM}}(z)$.

Let $w(z) := \exp_z^{-1}(\overline{z}) \in T_z\widetilde{\mathcal{M}}_{p,q}$. By the smoothness of the exponential map $\exp : \mathcal{T}\widetilde{\mathcal{M}}_{p,q} \to \widetilde{\mathcal{M}}_{p,q}$, we have

$$w(z) = -(z - \overline{z}) + O(\|z - \overline{z}\|^2),$$

and hence $z - \overline{z} = -w(z) + O(\|z - \overline{z}\|^2)$. Furthermore, since $\|v\| = O(\|z - \overline{z}\|)$ (implied by the boundedness of the operator inverse and smoothness of F), we have

$$\begin{aligned}
&\left\|R_z(v) - \overline{z}\right\|\\
&= \left\|\left((dF_z)^* dF_z + \mu(z)I\right)^{-1}\Big[(dF_z)^*\big(dF_z w(z) + F(z)\big) + \mu(z)w(z)\Big]\right\| + O(\|z - \overline{z}\|^2)\\
&\le \left\|\left((dF_z)^* dF_z + \mu(z)I\right)^{-1}\right\|\Big(\|(dF_z)^*\big(dF_z w(z) + F(z)\big)\| + \|\mu(z)w(z)\|\Big) + O(\|z - \overline{z}\|^2).
\end{aligned}$$

We now bound each component:

(i) By Corollary 2, the W-SOC and the W-SRCQ at $\overline{z}$ imply the uniform boundedness

$$\left\| \left((dF_z)^* dF_z + \mu(z) I \right)^{-1} \right\| = O(1).$$

(ii) Because $w(z) = \exp_z^{-1}(\overline{z})$, the first-order Taylor expansion of F along the geodesic from z to $\overline{z}$ gives

$$F(z) + dF_z w(z) = O(\|z - \overline{z}\|^2),$$

where the $O(\|z - \overline{z}\|^2)$ term is uniform still by the smoothness of exponential map and F. Hence, by the uniform boundedness of dF_z implied by the smoothness of F on $\widetilde{\mathcal{M}}_{p,q}$, we have $\|(dF_z)^*(dF_z w(z) + F(z))\| = O(\|z - \overline{z}\|^2)$.

(iii) By definition, $\mu(z) = \|F(z)\|^2 = O(\|z - \overline{z}\|^2)$, and $\|w(z)\| = O(\|z - \overline{z}\|)$. Thus,

$$\|\mu(z) w(z)\| = O(\|z - \overline{z}\|^3).$$

Combining estimates (i)–(iii), we finally conclude that

$$\left\| R_z(v) - \overline{z} \right\| = O(\|z - \overline{z}\|^2),$$

which completes the proof. □

Theorem 10 (restated). *Suppose z^∞ is an accumulation point of the sequence generated by Algorithm 1. If $0 < \delta < \delta(z^\infty)$ and z^∞ is a KKT pair at which the W-SOC and the SRCQ hold, then the whole sequence converges to z^∞ quadratically in the sense that*

$$\|z^{k+1} - z^\infty\| \leq O(\|z^k - z^\infty\|^2).$$

In particular, the sequence eventually lies on the stratum containing z^∞.

Proof z^∞ is a D-stationary point by Theorem 9. If the additional conditions hold, let $\{z^k\}_{k\in\theta}$ denote the subsequence converging to z^∞. Then, by the continuity of eigenvalues and $0 < \delta < \delta(z^\infty)$, the correction point $\widehat{z^k}$ generated by z^k (see (67)) lies on the stratum containing z^∞ for large enough $k \in \theta$. By Lemma 6, as $z^k \to_{k\in\theta} z^\infty$, we know $\widehat{z^k} \to_{k\in\theta} z^\infty$.

Now, consider any z lies on the stratum containing z^∞. By Proposition 11, once z sufficiently close to z^∞

$$\varphi(R_z(v)) - \varphi(z) = \frac{1}{2}\varphi'(z, v) + O(\|v\|^2), \tag{145}$$

where $v = v^{\mathrm{LM}}(z)$ is given by (65). Using the fact that $\varphi'(z, v)$ is of the same order as $\|v\|$, which is implied by Lemma 9, we conclude that the line search of the stratum LM step starting from z takes the unit step. Combining this with Proposition 12, which shows

$$\|R_z(v) - z^\infty\| \leq O(\|z - z^\infty\|^2), \tag{146}$$

we obtain

$$\varphi(R_z(v)) \leq O(\|R_z(v) - z^\infty\|^2) \leq O(\|z - z^\infty\|^4) \leq O(\varphi(z)^2), \tag{147}$$

where the first inequality follows from the Lipschitz continuity of F, and the last inequality follows from the stratum-restricted local error bound (36).

Now, for $k \in \theta$ large enough, denote by z_1^{k+1} and z_2^{k+1} the normal steps from $\widehat{z^k}$ (see Algorithm 2). Then, by Proposition 10, we have

$$\varphi(\widehat{z^k}) \leq c_3 \min\{\varphi(z^k), \varphi(z_1^{k+1}), \varphi(z_2^{k+1})\}, \tag{148}$$

which, together with (147) and the fact that $\varphi(\widehat{z^k}) \to 0$, implies

$$\varphi(R_{\widehat{z^k}}(\widehat{v}^k)) < \min\{\varphi(z^k), \varphi(z_1^{k+1}), \varphi(z_2^{k+1})\}, \tag{149}$$

where $\widehat{v^k} = v^{\mathrm{LM}}(\widehat{z^k})$. Finally, by the logic of Algorithm 1, we know that the next iterate must be $z^{k+1} = R_{\widehat{z^k}}(\widehat{v}^k)$. Moreover, we have $\widehat{z^{k+1}} = z^{k+1}$ and, by (146), $\|z^{k+1} - z^\infty\| < \|z^k - z^\infty\|$.

By a recursive argument, we can conclude that, for k large enough, z^k lies on the stratum containing z^∞, $z^{k+1} = R_{z^k}(v^{\mathrm{LM}}(z^k))$, and $z^k \to z^\infty$ at a quadratic rate. □