

1 EUCLIDEAN REPRESENTATION OF LOW-RANK MATRICES AND 2 ITS GEOMETRIC PROPERTIES *

3 FANGZHENG XIE[†]

4 **Abstract.** In this paper, we propose a novel and user-friendly Euclidean representation frame-
5 work for low-rank matrices. Correspondingly, we establish a collection of technical and theoretical
6 tools for analyzing the intrinsic perturbation and geometry of low-rank matrices in which the under-
7 lying referential matrix and the perturbed matrix both live on the same low-rank matrix manifold. It
8 turns out that the Frobenius distance between low-rank matrices is locally equivalent to the Euclid-
9 ean distance between the corresponding representing vectors. Our analyses also show that, locally
10 around the referential matrix, the sine-theta distance between subspaces is equivalent to the Euclid-
11 ean distance between two appropriately selected orthonormal basis, circumventing the orthogonal
12 Procrustes analysis. We further establish the regularity of the proposed Euclidean representation
13 function by showing that it has a non-degenerate Fréchet derivative, thereby establishing the manifold
14 structure over the collection of all low-rank matrices. Two applications of the proposed framework are
15 showcased, namely, constructing the Riemannian structure on the Grassmannian using the Cayley
16 parameterization and the multiplicative perturbation analysis of low-rank matrices.

17 **Key words.** Cayley parameterization, Euclidean representation, intrinsic perturbation analysis,
18 low-rank matrix geometry, low-rank matrix manifold

19 **AMS subject classifications.** 15A18, 15A42, 15B10

20 **1. Introduction.** Due to the emergence of high-dimensional data, low-rank ma-
21 trix models have been extensively studied and broadly applied in statistics, proba-
22 bility, machine learning, optimization, applied mathematics, and various application
23 domains. For example, principal component analysis [6, 12, 36, 41, 54, 55, 65], co-
24 variance matrix estimation [11, 13, 24, 29, 35], matrix completion [14, 15, 16, 20, 23],
25 random graph inference [1, 7, 47, 50, 51, 60], among others, heavily rely on low-rank
26 matrix models. From the practical perspective, specific domain-oriented applications
27 involving low-rank matrices include collaborative filtering [32], neural science [27, 52],
28 social networks [44, 61], and cryo-EM [48].

29 Spectral methods have been ubiquitous to gain insight into low-rank matrix mod-
30 els in the presence of high-dimensional data. For example, in the stochastic block
31 model, the K -means clustering procedure is applied to the rows of the leading eigen-
32 vector matrix of the observed adjacency matrix or its normalized Laplacian matrix
33 to discover the underlying community structure [1, 47, 50, 51]. Meanwhile, the the-
34oretical understanding of spectral methods has also been developed based on matrix
35 perturbation analysis [9, 14, 21, 49, 56, 62]. Specifically, given an approximately low-
36 rank matrix Σ_0 and a perturbation matrix \mathbf{E} that is comparatively smaller than Σ_0
37 in magnitude, matrix perturbation analysis studies how the eigenspaces or singular
38 subspaces of the perturbed matrix $\Sigma := \Sigma_0 + \mathbf{E}$ differ from those of the original
39 matrix Σ_0 in terms of the behavior of the perturbation \mathbf{E} . Notably, when Σ and Σ_0
40 are symmetric matrices, the famous Davis-Kahan theorem [21, 62] asserts that the
41 distance between the subspace spanned by the leading eigenvectors of Σ and that of
42 Σ_0 , formally defined via the canonical angles (see Section 2.1 below), can be upper
43 bounded by the matrix norms of \mathbf{E} . Several extensions and generalizations of the ma-
44 trix perturbation tools have been developed. In [14, 56], the authors later extended
45 the Davis-Kahan theorem to deal with singular subspaces and rectangular matrices.

*Submitted to the editors DATE.

[†]Department of Statistics, Indiana University, Bloomington, IN, 47408 (fxie@iu.edu).

In the context where \mathbf{E} is a mean-zero random matrix, the authors of [45, 53] obtained optimal and sharp results that improve the classical deterministic bounds due to [21, 56]. Recently, there has also been a collection of works focusing on the entrywise perturbation behavior of eigenvectors as well as the two-to-infinity norm of eigenvector perturbation analysis when \mathbf{E} is deterministic or random [2, 17, 18, 28, 30].

This work is motivated by the intrinsic perturbation analysis of low-rank matrices, in which both the referential matrix Σ_0 and the perturbed matrix $\Sigma = \Sigma_0 + \mathbf{E}$ lie on the same low-rank matrix manifold. A key feature of this setup is that the rank of Σ is the same as the rank of Σ_0 . This is slightly different from the classical random perturbation setup where the referential matrix Σ_0 is perturbed by a mean-zero but potentially full-rank random matrix \mathbf{E} , as the resulting perturbed matrix Σ may not necessarily be low-rank. Nevertheless, understanding the intrinsic perturbation of low-rank matrices is of fundamental interest in many applications, such as the optimization algorithms with orthogonality constraints [26], sampling random orthogonal matrices [34], multiplicative perturbation [38, 39]. In the context of intrinsically perturbed low-rank matrices, the classical tools following the Davis-Kahan framework (e.g., those developed in [21, 56, 62]), although still valid, are less user-friendly to obtain sharp and optimal results in various problems. This is largely because that, in scenarios where the analysis of the difference between two matrices having the same rank is desired, the perturbation matrix \mathbf{E} is structurally more complicated for analysis. Hence, it brings additional technical challenges when the Davis-Kahan framework is applied directly.

This paper establishes a novel Euclidean representation framework for low-rank matrices and provides a collection of theoretical and technical tools for studying the intrinsic perturbation and geometric properties of low-rank matrices. Specifically, leveraging the Cayley parameterization for subspaces [34], we propose a matrix-valued function to represent generic low-rank matrices using vectors in an open subset of the Euclidean space. Furthermore, built upon the proposed Euclidean representation framework, we show that the intrinsic perturbation of low-rank matrices can be characterized by the behavior of their representing Euclidean vectors. Consequently, the Frobenius sine-theta distance between subspaces (formally defined in Section 2 below) is locally equivalent to the Frobenius distance between two suitably selected Stiefel matrices spanning the corresponding subspaces, which is user-friendly and circumvents the need for an orthogonal Procrustes analysis. Another fundamental result of the proposed framework is that the collection of low-rank matrices of interest can be viewed as a Euclidean manifold, and our proposed Euclidean representation function directly yields an atlas for the low-rank matrix manifold, thereby providing a gateway to further geometric structures such as the tangent spaces, Riemannian metric, connections, and geodesics on the low-rank matrix manifold. Two applications of the proposed framework are presented: the construction of the Riemannian structure on the Grassmannian using the Cayley parameterization and the multiplicative perturbation analysis of low-rank matrices.

The rest of the paper is organized as follows. In Section 2, we present the proposed Euclidean representation framework for low-rank matrices after the introduction of basic notations and definitions. Section 3 elaborates on our main technical results, including the intrinsic perturbation theorems and the regularity theorem of the proposed Euclidean representation function. Section 4 showcases two applications of the proposed framework: the construction of the Riemannian structure on the Grassmannian using the Cayley parameterization and the multiplicative perturbation analysis of low-rank matrices. In Section 5, we extend the current framework

to general rectangular matrices. Further discussion is included in Section 6. The proofs of the technical results are contained in Appendix.

2. Preliminaries.

2.1. Notations and definitions. We use the symbol $:=$ to assign mathematical definitions of quantities. The notation \mathbf{A}^\dagger denotes the Moore-Penrose pseudoinverse of an arbitrary matrix \mathbf{A} . For any two positive semidefinite matrices \mathbf{A} and \mathbf{B} of the same dimension, we use $\mathbf{A} \succeq \mathbf{B}$ ($\mathbf{A} \preceq \mathbf{B}$, resp.) to indicate that $\mathbf{A} - \mathbf{B}$ is positive semidefinite ($\mathbf{B} - \mathbf{A}$ is positive semidefinite, resp.). The $r \times r$ identity matrix is denoted by \mathbf{I}_r , the p -dimensional zero vector is denoted by $\mathbf{0}_p$, and the $p \times q$ zero matrix is denoted by $\mathbf{0}_{p \times q}$. Given two positive integers p, r , we denote $\mathbb{O}(p, r) = \{\mathbf{U} \in \mathbb{R}^{p \times r} : \mathbf{U}^T \mathbf{U} = \mathbf{I}_r\}$ the collection of all $p \times r$ Stiefel matrices and write $\mathbb{O}(r) := \mathbb{O}(r, r)$. For any $\mathbf{Y} \in \mathbb{R}^{p \times r}$, we use $\text{Span}(\mathbf{Y})$ to denote the r -dimensional subspace in \mathbb{R}^p spanned by the columns of \mathbf{Y} . The collection of all $r \times r$ symmetric matrices is denoted by $\mathbb{M}(r)$ and the collection of all $r \times r$ symmetric positive definite matrices is denoted by $\mathbb{M}_+(r)$. For a matrix $\mathbf{\Sigma} \in \mathbb{R}^{p_1 \times p_2}$ and indices $i \in [p_1], j \in [p_2]$, let $[\mathbf{\Sigma}]_{ij}$ denote the element on the i th row and j th column of $\mathbf{\Sigma}$, $[\mathbf{\Sigma}]_{i*}$ denote the i th row of $\mathbf{\Sigma}$, and $[\mathbf{\Sigma}]_{j*}$ denote the j th column of $\mathbf{\Sigma}$. Furthermore, we use $\sigma_1(\mathbf{\Sigma}), \dots, \sigma_{p_1 \wedge p_2}(\mathbf{\Sigma})$ to denote the singular values of $\mathbf{\Sigma}$ sorted in the non-increasing order, i.e., $\sigma_1(\mathbf{\Sigma}) \geq \dots \geq \sigma_{p_1 \wedge p_2}(\mathbf{\Sigma})$. When $\mathbf{\Sigma}$ is a $p \times p$ symmetric square matrix, $\lambda_1(\mathbf{\Sigma}), \dots, \lambda_p(\mathbf{\Sigma})$ denote the eigenvalues of $\mathbf{\Sigma}$ sorted in the non-increasing order in magnitude, namely, $|\lambda_1(\mathbf{\Sigma})| \geq \dots \geq |\lambda_p(\mathbf{\Sigma})|$. The spectral norm of a general matrix $\mathbf{\Sigma}$, denoted by $\|\mathbf{\Sigma}\|_2$, is the largest singular value of $\mathbf{\Sigma}$, and the Frobenius norm of $\mathbf{\Sigma}$, denoted by $\|\mathbf{\Sigma}\|_F$, is defined to be $\|\mathbf{\Sigma}\|_F = (\sum_{i=1}^{p_1} \sum_{j=1}^{p_2} [\mathbf{\Sigma}]_{ij}^2)^{1/2}$. For a Euclidean vector $\mathbf{x} = [x_1, \dots, x_p]^T \in \mathbb{R}^p$, we denote $[\mathbf{x}]_i := x_i$, $\|\mathbf{x}\|_2$ the usual Euclidean norm $\|\mathbf{x}\|_2 = (\sum_i x_i^2)^{1/2}$, let $B_2(\mathbf{x}, \epsilon) := \{\mathbf{y} \in \mathbb{R}^p : \|\mathbf{y} - \mathbf{x}\|_2 < \epsilon\}$, and let $\text{diag}(\mathbf{x})$ be the $p \times p$ diagonal matrix with x_i being the element on its i th row and i th column.

For a $p_1 \times p_2$ matrix \mathbf{M} , the operator $\text{vec}(\cdot)$ transform \mathbf{M} to a $p_1 p_2$ -dimensional Euclidean vector by stacking the columns of $\mathbf{\Sigma}$ consecutively, i.e.,

$$\text{vec}(\mathbf{M}) = [[\mathbf{M}]_{11}, \dots, [\mathbf{M}]_{p_1 1}, [\mathbf{M}]_{12}, \dots, [\mathbf{M}]_{p_1 2}, \dots, [\mathbf{M}]_{1 p_2}, \dots, [\mathbf{M}]_{p_1 p_2}]^T.$$

The operator $\text{vech}(\cdot)$ transform an $r \times r$ square symmetric matrix \mathbf{M} to an $r(r+1)/2$ -dimensional Euclidean vector by eliminating all its super-diagonal elements, i.e.,

$$\text{vech}(\mathbf{M}) = [[\mathbf{M}]_{11}, [\mathbf{M}]_{21}, \dots, [\mathbf{M}]_{r1}, [\mathbf{M}]_{22}, \dots, [\mathbf{M}]_{r2}, \dots, [\mathbf{M}]_{rr}]^T.$$

For any two positive integers p, q , we denote \mathbf{K}_{pq} the $pq \times pq$ commutation matrix such that $\text{vec}(\mathbf{M}^T) = \mathbf{K}_{pq} \text{vec}(\mathbf{M})$ for any $\mathbf{M} \in \mathbb{R}^{p \times q}$, and denote \mathbb{D}_p the duplication matrix such that $\text{vec}(\mathbf{M}) = \mathbb{D}_p \text{vech}(\mathbf{M})$ for any symmetric $\mathbf{M} \in \mathbb{R}^{p \times p}$. We refer the readers to [42] for a review of the properties of the commutation matrix \mathbf{K}_{pq} and the duplication matrix \mathbb{D}_p . For two matrices $\mathbf{A} \in \mathbb{R}^{p \times q}$ and $\mathbf{B} \in \mathbb{R}^{m \times n}$, we use $\mathbf{A} \otimes \mathbf{B}$ to denote the Kronecker product of \mathbf{A} and \mathbf{B} .

The distance between linear subspaces can be measured in terms of the canonical angles, formally defined as follows. Given two Stiefel matrices $\mathbf{U}, \mathbf{U}_0 \in \mathbb{O}(p, r)$, let $\sigma_1(\mathbf{U}_0^T \mathbf{U}) \geq \dots \geq \sigma_r(\mathbf{U}_0^T \mathbf{U}) \geq 0$ be the singular values of $\mathbf{U}_0^T \mathbf{U}$. Note the singular values of $\mathbf{U}_0^T \mathbf{U}$ are unitarily invariant and only depend on $\text{Span}(\mathbf{U})$ and $\text{Span}(\mathbf{U}_0)$. The canonical angles between \mathbf{U}_0 and \mathbf{U} are defined to be the diagonal entries of $\Theta(\mathbf{U}_0, \mathbf{U}) := \text{diag}[\cos^{-1}\{\sigma_1(\mathbf{U}_0^T \mathbf{U})\}, \dots, \cos^{-1}\{\sigma_r(\mathbf{U}_0^T \mathbf{U})\}] \in \mathbb{R}^{r \times r}$. Then the spectral sine-theta distance and the Frobenius sine-theta distance between $\text{Span}(\mathbf{U}_0)$ and $\text{Span}(\mathbf{U})$ are defined by $\|\sin \Theta(\mathbf{U}_0, \mathbf{U})\|_2$ and $\|\sin \Theta(\mathbf{U}_0, \mathbf{U})\|_F$, respectively.

2.2. Euclidean representation of subspaces. We first introduce the Cayley parameterization of Subspaces proposed by [34], which serves as an intermediate step towards our proposed Euclidean representation framework for low-rank matrices.

The collection of all r -dimensional subspaces in \mathbb{R}^p is of fundamental interest in multivariate statistics. When equipped with an appropriate topology and an atlas, the collection of all r -dimensional subspace in \mathbb{R}^p is referred to as the Grassmannian and is denoted by $\mathcal{G}(p, r)$. Nevertheless, the elements in $\mathcal{G}(p, r)$ are too abstract and inconvenient for analysis. It is therefore desirable to find a concrete and user-friendly representation of general r -dimensional subspace in \mathbb{R}^p .

Suppose $\mathbb{S} \subset \mathbb{R}^p$ is an r -dimensional linear subspace in \mathbb{R}^p . It is always possible to find a Stiefel matrix $\mathbf{U} \in \mathbb{O}(p, r)$ whose columns span \mathbb{S} , and one may use \mathbf{U} as a representer for the subspace \mathbb{S} . The disadvantage of this representation is that \mathbf{U} cannot be uniquely identified by \mathbb{S} , since for any orthogonal rotation matrix $\mathbf{W} \in \mathbb{O}(r)$, $\text{Span}(\mathbf{U}) = \text{Span}(\mathbf{U}\mathbf{W})$. Such non-identifiability of orthonormal basis brings natural inconvenience to, for example, statistical analyses because the Fisher information matrix with a non-identifiable parameterization will be singular. Thanks to the result of [34], almost every r -dimensional subspace in \mathbb{R}^p can be uniquely represented by a Stiefel matrix $\mathbf{U} = [\mathbf{Q}_1^T, \mathbf{Q}_2^T]^T$ in $\mathbb{O}(p, r)$ such that $\mathbf{Q}_1 \in \mathbb{R}^{r \times r}$ is symmetric positive definite with respect to the uniform probability measure of $\mathcal{G}(p, r)$. Formally, denote

$$\mathbb{O}_+(p, r) = \left\{ \mathbf{U} = \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix} \in \mathbb{O}(p, r) : \mathbf{Q}_1 \in \mathbb{M}_+(r) \right\}$$

and suppose $l : \mathbb{O}_+(p, r) \rightarrow \mathcal{G}(p, r)$ is the map defined by $\ell(\mathbf{U}) = \text{Span}(\mathbf{U})$. By Proposition 3.2 in [34], the image of l has probability one with respect to the uniform probability distribution on $\mathcal{G}(p, r)$. Hence, with (uniform) probability one, every r -dimensional subspace in \mathbb{R}^p can be uniquely represented by a Stiefel matrix in $\mathbb{O}_+(p, r)$, and therefore, finding a suitable representation of subspaces reduces to finding a suitable representation of Stiefel matrices in $\mathbb{O}_+(p, r)$. The case where the subspace \mathbb{S} cannot be represented by a $\mathbf{U} \in \mathbb{O}_+(p, r)$ will be addressed using a standard patching argument in differential manifold theory (see Section 4.1 for details).

Let \mathbf{A} be a $(p-r) \times r$ matrix with $\|\mathbf{A}\|_2 < 1$, and denote $\boldsymbol{\varphi} = \text{vec}(\mathbf{A})$. Then for any $\mathbf{U} \in \mathbb{O}_+(p, r)$, the Cayley parameterization of \mathbf{U} via $\boldsymbol{\varphi} = \text{vec}(\mathbf{A})$ is defined as the following map [34]:

$$(2.1) \quad \mathbf{U} : \boldsymbol{\varphi} \in \mathbb{R}^{(p-r)r} \mapsto \mathbf{U}(\boldsymbol{\varphi}) := (\mathbf{I}_p + \mathbf{X}_{\boldsymbol{\varphi}})(\mathbf{I}_p - \mathbf{X}_{\boldsymbol{\varphi}})^{-1} \mathbf{I}_{p \times r} \in \mathbb{O}(p, r),$$

where

$$(2.2) \quad \mathbf{X}_{\boldsymbol{\varphi}} = \begin{bmatrix} r \times r & -\mathbf{A}^T \\ \mathbf{A} & (p-r) \times (p-r) \end{bmatrix}, \quad \text{and} \quad \mathbf{I}_{p \times r} = \begin{bmatrix} \mathbf{I}_r \\ (p-r) \times r \end{bmatrix}.$$

The Cayley parameterization (2.1) above immediately leads to the following explicit expression for the submatrices of \mathbf{U} :

$$\mathbf{U} = \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix} = \begin{bmatrix} (\mathbf{I}_r - \mathbf{A}^T \mathbf{A})(\mathbf{I}_r + \mathbf{A}^T \mathbf{A})^{-1} \\ 2\mathbf{A}(\mathbf{I}_r + \mathbf{A}^T \mathbf{A})^{-1} \end{bmatrix}.$$

By Proposition 3.4 in [34], the Cayley parameterization $\mathbf{U}(\cdot)$, viewed as a map from $\{\boldsymbol{\varphi} = \text{vec}(\mathbf{A}) : \mathbf{A} \in \mathbb{R}^{(p-r) \times r}, \|\mathbf{A}\|_2 < 1\}$ to $\mathbb{R}^{p \times r}$, is also differentiable with the Fréchet derivative

$$(2.3) \quad D\mathbf{U}(\boldsymbol{\varphi}) = 2[\mathbf{I}_{p \times r}^T (\mathbf{I}_p - \mathbf{X}_{\boldsymbol{\varphi}})^{-T} \otimes (\mathbf{I}_p - \mathbf{X}_{\boldsymbol{\varphi}})^{-1}] \boldsymbol{\Gamma}_{\boldsymbol{\varphi}},$$

where $\mathbf{\Gamma}_\varphi$ is a matrix such that $\mathbf{\Gamma}_\varphi \varphi = \text{vec}(\mathbf{X}_\varphi)$. An explicit formula for $\mathbf{\Gamma}_\varphi$ is also available [34]: $\mathbf{\Gamma}_\varphi = (\mathbf{I}_{p^2} - \mathbf{K}_{pp})(\mathbf{\Theta}_1^T \otimes \mathbf{\Theta}_2^T)$, where $\mathbf{\Theta}_1 = \mathbf{I}_{p \times r}^T$, and $\mathbf{\Theta}_2 = [_{(p-r) \times r}, \mathbf{I}_{p-r}]$. The following theorem is a refined version of Proposition 3.4 in [34] in terms of a global and dimension-free control of the remainder.

THEOREM 2.1. *Let $\mathbf{U} : \varphi \mapsto \mathbf{U}(\varphi)$ be the Cayley parameterization defined by (2.1). Denote $D\mathbf{U}(\varphi)$ the Fréchet derivative of \mathbf{U} given by (2.3). Then for all φ, φ_0 , there exists a matrix $\mathbf{R}_\mathbf{U}(\varphi, \varphi_0) \in \mathbb{R}^{p \times r}$, such that*

$$\mathbf{U}(\varphi) - \mathbf{U}(\varphi_0) = 2(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}(\mathbf{X}_\varphi - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\mathbf{I}_{p \times r} + \mathbf{R}_\mathbf{U}(\varphi, \varphi_0),$$

where the remainder $\mathbf{R}_\mathbf{U}$ satisfies

$$\|\mathbf{R}_\mathbf{U}(\varphi, \varphi_0)\|_F \leq 8\|\varphi - \varphi_0\|_2^2.$$

In particular, $\|\mathbf{U}(\varphi) - \mathbf{U}(\varphi_0)\|_F \leq 2\sqrt{2}\|\varphi - \varphi_0\|_2$ for all φ and φ_0 .

Proof. First observe that by definition of $\mathbf{\Gamma}_\varphi$, we have,

$$\|\mathbf{\Gamma}_\varphi\|_2 = \max_{\|\varphi\|_2=1} \|\mathbf{\Gamma}_\varphi \varphi\|_2 = \max_{\|\varphi\|_2=1} \|\text{vec}(\mathbf{X}_\varphi)\|_2 = \max_{\|\varphi\|_2=1} (2\|\text{vec}(\mathbf{A})\|_2^2)^{1/2} = \sqrt{2}.$$

Because $\mathbf{X}_\varphi^T = -\mathbf{X}_\varphi$, we also have

$$\begin{aligned} \|(\mathbf{I}_p + \mathbf{X}_\varphi)^{-1}\|_2 &= \|(\mathbf{I}_p - \mathbf{X}_\varphi)^{-1}\|_2 = \lambda_{\min}^{-1/2}((\mathbf{I}_p + \mathbf{X}_\varphi)(\mathbf{I}_p - \mathbf{X}_\varphi)) \\ &= \lambda_{\min}^{-1/2}(\mathbf{I}_p + \mathbf{X}_\varphi \mathbf{X}_\varphi^T) \leq \lambda_{\min}^{-1/2}(\mathbf{I}_p) = 1 \end{aligned}$$

for any φ . Write

$$\begin{aligned} \|(\mathbf{I}_p - \mathbf{X}_\varphi)^{-1} - (\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\|_F &= \|(\mathbf{I}_p - \mathbf{X}_\varphi)^{-1}(\mathbf{X}_\varphi - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\|_F \\ &\leq \|(\mathbf{I}_p - \mathbf{X}_\varphi)^{-1}\|_2 \|\mathbf{X}_\varphi - \mathbf{X}_{\varphi_0}\|_F \|(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\|_2 \\ &\leq \|\mathbf{X}_\varphi - \mathbf{X}_{\varphi_0}\|_F. \end{aligned}$$

Furthermore, by matrix algebra,

$$\begin{aligned} &(\mathbf{I}_p - \mathbf{X}_\varphi)^{-1} - (\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1} \\ &= (\mathbf{I}_p - \mathbf{X}_\varphi)^{-1}(\mathbf{X}_\varphi - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1} \\ &= (\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}(\mathbf{X}_\varphi - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1} \\ &\quad + \{(\mathbf{I}_p - \mathbf{X}_\varphi)^{-1} - (\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\}(\mathbf{X}_\varphi - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}. \end{aligned}$$

Denote $\mathbf{T}(\varphi, \varphi_0) := \{(\mathbf{I}_p - \mathbf{X}_\varphi)^{-1} - (\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\}(\mathbf{X}_\varphi - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}$. It follows that

$$(\mathbf{I}_p - \mathbf{X}_\varphi)^{-1} - (\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1} = (\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}(\mathbf{X}_\varphi - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1} + \mathbf{T}(\varphi, \varphi_0),$$

and

$$\|\mathbf{T}(\varphi, \varphi_0)\|_F \leq \|\mathbf{X}_\varphi - \mathbf{X}_{\varphi_0}\|_F^2.$$

Therefore, we can write

$$\begin{aligned}
\mathbf{U}(\varphi) - \mathbf{U}(\varphi_0) &= (\mathbf{I}_p + \mathbf{X}_{\varphi_0} + \mathbf{X}_{\varphi} - \mathbf{X}_{\varphi_0})\{(\mathbf{I}_p - \mathbf{X}_{\varphi})^{-1} - (\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\}\mathbf{I}_{p \times r} \\
&\quad + (\mathbf{X}_{\varphi} - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\mathbf{I}_{p \times r} \\
&= (\mathbf{I}_p + \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}(\mathbf{X}_{\varphi} - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\mathbf{I}_{p \times r} \\
&\quad + (\mathbf{I}_p + \mathbf{X}_{\varphi_0})\mathbf{T}(\varphi, \varphi_0)\mathbf{I}_{p \times r} \\
&\quad + (\mathbf{X}_{\varphi} - \mathbf{X}_{\varphi_0})\{(\mathbf{I}_p - \mathbf{X}_{\varphi})^{-1} - (\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\}\mathbf{I}_{p \times r} \\
&\quad + (\mathbf{I}_p - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}(\mathbf{X}_{\varphi} - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\mathbf{I}_{p \times r} \\
&= 2(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}(\mathbf{X}_{\varphi} - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\mathbf{I}_{p \times r} + \mathbf{R}_{\mathbf{U}}(\varphi, \varphi_0),
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{R}_{\mathbf{U}}(\varphi, \varphi_0) &:= (\mathbf{I}_p + \mathbf{X}_{\varphi_0})\mathbf{T}(\varphi, \varphi_0)\mathbf{I}_{p \times r} \\
&\quad + (\mathbf{X}_{\varphi} - \mathbf{X}_{\varphi_0})\{(\mathbf{I}_p - \mathbf{X}_{\varphi})^{-1} - (\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\}\mathbf{I}_{p \times r}.
\end{aligned}$$

Using the aforementioned results, we further compute

$$\begin{aligned}
\|\mathbf{R}_{\mathbf{U}}(\varphi, \varphi_0)\|_{\mathbf{F}} &\leq \|\mathbf{I}_p + \mathbf{X}_{\varphi_0}\|_2 \|\mathbf{T}(\varphi, \varphi_0)\|_{\mathbf{F}} + \|\mathbf{X}_{\varphi} - \mathbf{X}_{\varphi_0}\|_{\mathbf{F}}^2 \\
&\leq (1 + 2\|\mathbf{A}\|_2) \|\mathbf{X}_{\varphi} - \mathbf{X}_{\varphi_0}\|_{\mathbf{F}}^2 + \|\mathbf{X}_{\varphi} - \mathbf{X}_{\varphi_0}\|_{\mathbf{F}}^2 \\
&\leq 8\|\varphi - \varphi_0\|_2^2.
\end{aligned}$$

The proof is thus completed. \square

The authors of [34] also showed that the Cayley parameterization is one-to-one, and we refer to the inverse map $\mathbf{A} : \mathbb{O}_+(p, r) \rightarrow \{\mathbf{A} \in \mathbb{R}^{(p-r) \times r} : \|\mathbf{A}\|_2 < 1\}$ as the inverse Cayley parameterization. Formally, for any $\mathbf{U} = [\mathbf{Q}_1^{\mathbf{T}}, \mathbf{Q}_2^{\mathbf{T}}]^{\mathbf{T}} \in \mathbb{O}_+(p, r)$ where $\mathbf{Q}_1 \in \mathbb{M}_+(r)$, the inverse Cayley parameterization of \mathbf{U} is given by

$$(2.4) \quad \mathbf{A} = \mathbf{A}(\mathbf{U}) = \mathbf{Q}_2(\mathbf{I}_r + \mathbf{Q}_1)^{-1}.$$

The following theorem claims that the map $\mathbf{A}(\cdot)$ is globally Lipschitz continuous.

THEOREM 2.2. *Let $\mathbf{A}(\cdot)$ be the inverse Cayley parameterization defined by (2.4). Then for all $\mathbf{U}, \mathbf{U}_0 \in \mathbb{O}_+(p, r)$, $\|\mathbf{A}(\mathbf{U}) - \mathbf{A}(\mathbf{U}_0)\|_{\mathbf{F}} \leq 2\|\mathbf{U} - \mathbf{U}_0\|_{\mathbf{F}}$.*

Proof. Observe that

$$\begin{aligned}
&\|(\mathbf{I}_r + \mathbf{Q}_1)^{-1} - (\mathbf{I}_r + \mathbf{Q}_{01})^{-1}\|_{\mathbf{F}} \\
&= \|(\mathbf{I}_r + \mathbf{Q}_1)^{-1}\{(\mathbf{I}_r + \mathbf{Q}_{01}) - (\mathbf{I}_r + \mathbf{Q}_1)\}(\mathbf{I}_r + \mathbf{Q}_{01})^{-1}\|_{\mathbf{F}} \\
&\leq \|(\mathbf{I}_r + \mathbf{Q}_1)^{-1}\|_2 \|\mathbf{Q}_1 - \mathbf{Q}_{01}\|_{\mathbf{F}} \|(\mathbf{I}_r + \mathbf{Q}_{01})^{-1}\|_2 \\
&\leq \|\mathbf{Q}_1 - \mathbf{Q}_{01}\|_{\mathbf{F}}
\end{aligned}$$

because $\mathbf{I}_r + \mathbf{Q}_1 \succeq \mathbf{I}_r$ and $\mathbf{I}_r + \mathbf{Q}_{01} \succeq \mathbf{I}_r$. Therefore,

$$\begin{aligned}
&\|\mathbf{A}(\mathbf{U}) - \mathbf{A}(\mathbf{U}_0)\|_{\mathbf{F}} \\
&\leq \|\mathbf{Q}_2 - \mathbf{Q}_{02}\|_{\mathbf{F}} \|(\mathbf{I}_r + \mathbf{Q}_1)^{-1}\|_2 + \|\mathbf{Q}_{02}\|_2 \|(\mathbf{I}_r + \mathbf{Q}_1)^{-1} - (\mathbf{I}_r + \mathbf{Q}_{01})^{-1}\|_{\mathbf{F}} \\
&\leq \|\mathbf{Q}_2 - \mathbf{Q}_{02}\|_{\mathbf{F}} + \|\mathbf{Q}_1 - \mathbf{Q}_{01}\|_{\mathbf{F}} \\
&\leq 2\|\mathbf{U} - \mathbf{U}_0\|_{\mathbf{F}}.
\end{aligned}$$

The proof is thus completed. \square

2.3. Connection with other Euclidean representations of subspaces. It is clear that the Cayley parameterization (2.1) is not the unique way of representing subspaces in Grassmannian using Euclidean vectors. This subsection aims at establishing the connection between the Cayley parameterization and some classical subspace parameterization methods known in the literature.

In [3], the authors proposed the cross-section mapping for the Grassmannian $\mathcal{G}(p, r)$ as follows. For any $\text{Span}(\mathbf{Y}) \in \mathcal{G}(p, r)$ with $\mathbf{Y} \in \mathbb{R}^{p \times r}$, define

$$\mathcal{U}_{\mathbf{Y}} = \{\text{Span}(\mathbf{Z}) : \det(\mathbf{Z}^T \mathbf{Y}) \neq 0, \mathbf{Z} \in \mathbb{R}^{p \times r}\} \subset \mathcal{G}(p, r)$$

and the cross-section mapping is given by $\mathbf{B} \in \mathbb{R}^{(p-r) \times r} \mapsto \text{Span}(\mathbf{Y} + \mathbf{Y}_{\perp} \mathbf{B})$, where $\mathbf{Y}_{\perp} \in \mathbb{R}^{p \times (p-r)}$ spans the orthogonal complement of \mathbf{Y} . The rich Riemannian structures, including the tangent vector representation using the horizontal lift, the Riemannian metric, gradient, the Riemannian connection, and geodesics, are established using the cross-section mapping. In Section 4.1, we will show that the Cayley parameterization of subspaces can be applied to construct a Riemannian structure equivalent to the canonical one established in [3]. Nevertheless, the cross-section mapping is not advocated in the current work for the following two reasons: Conceptually, it is not convenient for the Euclidean parameterization of low-rank matrices; Practically, it is not directly applicable to numerical algorithms or practical computations because the subspaces are non-Euclidean objects.

It is more convenient to work with the Stiefel matrix representation of subspaces for the Euclidean parameterization of low-rank matrices and practical implementation using computer software. The above cross-section mapping gives rise to a choice of such a Stiefel matrix by taking the square root of the corresponding projection matrix:

$$(2.5) \quad \mathbf{B} \mapsto \mathbf{U}(\mathbf{I}_r + \mathbf{B}^T \mathbf{B})^{-1/2} + \mathbf{U}_{\perp} \mathbf{B}(\mathbf{I} + \mathbf{B}^T \mathbf{B})^{-1/2}.$$

See, for example, [4, 8, 25, 46, 57]. When $\mathbf{U} = \mathbf{P}\mathbf{I}_{p \times r}$ for some $p \times p$ permutation matrix \mathbf{P} , the above parameterization coincides with those constructed in [8] and [31]. In this case, up to a permutation matrix \mathbf{P} , the parameterization (2.5) has a one-to-one correspondence with the Cayley parameterization (2.1) through the relation $\mathbf{B} = 2\mathbf{A}(\mathbf{I}_r - \mathbf{A}^T \mathbf{A})^{-1}$.

2.4. Euclidean representation of low-rank matrices and intrinsic perturbation. We now leverage the aforementioned Cayley parameterization of subspaces and establish a Euclidean representation framework for symmetric low-rank matrices. Consider a symmetric $p \times p$ matrix $\mathbf{\Sigma}$ with $\text{rank}(\mathbf{\Sigma}) = r \leq p$. Let $\mathbf{\Sigma}$ yield the spectral decomposition $\mathbf{\Sigma} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$, where $\mathbf{V} \in \mathbb{O}(p, r)$ is the Stiefel matrix of eigenvectors, and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_r)$ is the diagonal matrix of non-zero eigenvalues of $\mathbf{\Sigma}$ with $|\lambda_1| \geq \dots \geq |\lambda_r| > 0$. In scenarios where the eigenvalues $\lambda_1, \dots, \lambda_r$ may include multiplicity, the eigenvector matrix \mathbf{V} may only be determined up to an orthogonal matrix in $\mathbb{O}(r)$. Also note by the aforementioned analysis, for almost every $\text{Span}(\mathbf{V}) \in \mathcal{G}(p, r)$ (with respect to the uniform probability measure over $\mathcal{G}(p, r)$), there exists another Stiefel matrix $\mathbf{U} \in \mathbb{O}_+(p, r)$ such that $\text{Span}(\mathbf{V}) = \text{Span}(\mathbf{U})$. It follows that almost every $\mathbf{\Sigma}$ with $\text{rank}(\mathbf{\Sigma}) = r$ can be reparameterized as $\mathbf{\Sigma} = \mathbf{U}\mathbf{M}\mathbf{U}^T$ for a $p \times r$ Stiefel matrix $\mathbf{U} \in \mathbb{O}_+(p, r)$ and a $r \times r$ symmetric matrix $\mathbf{M} \in \mathbb{M}(r)$, where the term “almost every” is with respect to the product measure of the uniform probability measure over $\mathbb{O}(p, r)$ and the Lebesgue measure over $\mathbb{M}(r)$. Then by the result of Section 2.2, there exists a unique $\mathbf{A} \in \mathbb{R}^{(p-r) \times r}$, $\|\mathbf{A}\|_2 < 1$, and $\varphi = \text{vec}(\mathbf{A})$, such that $\mathbf{U} = \mathbf{U}(\varphi)$, where $\mathbf{U}(\cdot)$ is the Cayley parameterization defined

by (2.1). Let $\boldsymbol{\mu} = \text{vech}(\mathbf{M})$. Conversely, the matrix \mathbf{M} can be viewed as a function of $\boldsymbol{\mu}$, denoted generically by $\mathbf{M}(\cdot) : \mathbb{R}^{r(r+1)/2} \rightarrow \mathbb{M}(r)$, as the inverse of the map $\mathbf{M} \mapsto \boldsymbol{\mu} = \text{vech}(\mathbf{M})$. Therefore, by denoting $\boldsymbol{\theta} = [\boldsymbol{\varphi}^T, \boldsymbol{\mu}^T]^T$, we can represent almost every $\boldsymbol{\Sigma}$ with $\text{rank}(\boldsymbol{\Sigma}) = r$ through the following matrix-valued function, which is generically denoted by $\boldsymbol{\Sigma}(\cdot)$:

$$(2.6) \quad \boldsymbol{\Sigma}(\cdot) : \mathcal{D}(p, r) \rightarrow \mathcal{S}(p, r), \quad \boldsymbol{\theta} \mapsto \mathbf{U}(\boldsymbol{\varphi})\mathbf{M}(\boldsymbol{\mu})\mathbf{U}(\boldsymbol{\varphi})^T,$$

where

$$(2.7) \quad \mathcal{D}(p, r) := \left\{ \boldsymbol{\theta} = \begin{bmatrix} \text{vec}(\mathbf{A}) \\ \boldsymbol{\mu} \end{bmatrix} \in \mathbb{R}^{(p-r)r} \times \mathbb{R}^{r(r+1)/2} : \mathbf{A} \in \mathbb{R}^{(p-r) \times r}, \|\mathbf{A}\|_2 < 1 \right\},$$

is the domain of the map $\boldsymbol{\Sigma}(\cdot)$ and

$$(2.8) \quad \mathcal{S}(p, r) := \{\boldsymbol{\Sigma} = \mathbf{U}\mathbf{M}\mathbf{U}^T : \mathbf{U} \in \mathbb{O}_+(p, r), \mathbf{M} \in \mathbb{M}(r)\}$$

is the collection of $p \times p$ rank- r matrices of interest. It is also clear that the domain $\mathcal{D}(p, r)$ for the parameter $\boldsymbol{\theta}$ is convex and connected. In fact, given any $\boldsymbol{\theta}_1 = [\text{vec}(\mathbf{A}_1)^T, \boldsymbol{\mu}_1^T]^T, \boldsymbol{\theta}_2 = [\text{vec}(\mathbf{A}_2)^T, \boldsymbol{\mu}_2^T]^T \in \mathcal{D}(p, r)$, where $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{(p-r) \times r}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^{r(r+1)/2}$ with $\|\mathbf{A}_1\|_2 < 1$ and $\|\mathbf{A}_2\|_2 < 1$, we have, for any $\lambda \in (0, 1)$,

$$(1 - \lambda)\boldsymbol{\theta}_1 + \lambda\boldsymbol{\theta}_2 = \begin{bmatrix} (1 - \lambda)\text{vec}(\mathbf{A}_1) + \lambda\text{vec}(\mathbf{A}_2) \\ (1 - \lambda)\boldsymbol{\mu}_1 + \lambda\boldsymbol{\mu}_2 \end{bmatrix} = \begin{bmatrix} \text{vec}\{(1 - \lambda)\mathbf{A}_1 + \lambda\mathbf{A}_2\} \\ (1 - \lambda)\boldsymbol{\mu}_1 + \lambda\boldsymbol{\mu}_2 \end{bmatrix}.$$

Since $\|(1 - \lambda)\mathbf{A}_1 + \lambda\mathbf{A}_2\|_2 \leq (1 - \lambda)\|\mathbf{A}_1\|_2 + \lambda\|\mathbf{A}_2\|_2 < (1 - \lambda) + \lambda = 1$, it follows that $(1 - \lambda)\boldsymbol{\theta}_1 + \lambda\boldsymbol{\theta}_2 \in \mathcal{D}(p, r)$, and hence, $\mathcal{D}(p, r)$ is convex and connected.

This paper is primarily interested in the intrinsic perturbation analysis between $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}$ with $\boldsymbol{\Sigma}, \boldsymbol{\Sigma}_0 \in \mathcal{S}(p, r)$, and $\mathbf{E} := \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0$ is assumed to be comparatively smaller than $\boldsymbol{\Sigma}_0$ in magnitude. In many statistical problems, $\boldsymbol{\Sigma}_0$ is the referential matrix of interest, but only the perturbed version $\boldsymbol{\Sigma}$ is accessible. To be more specific, $\boldsymbol{\Sigma}$ typically plays the role of a function of the observed data, namely, an estimator for the unknown $\boldsymbol{\Sigma}_0$. By the aforementioned analysis, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_0$ can be represented by some Euclidean vectors $\boldsymbol{\theta}, \boldsymbol{\theta}_0 \in \mathcal{D}(p, r)$, such that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. In turn, the problem of estimating the unobserved referential matrix $\boldsymbol{\Sigma}_0$ reduces to estimating the Euclidean representer $\boldsymbol{\theta}_0$ by an estimator $\boldsymbol{\theta}$, and hence, the perturbation analysis of $\boldsymbol{\Sigma}$ naturally translates to the problem of the perturbation analysis of $\boldsymbol{\theta}$.

We conclude this section with the introduction of the following matrix-valued functions. For any $\boldsymbol{\theta} = [\boldsymbol{\varphi}^T, \boldsymbol{\mu}^T]^T \in \mathcal{D}(p, r)$, let

$$(2.9) \quad \begin{aligned} D_{\boldsymbol{\varphi}}\boldsymbol{\Sigma}(\boldsymbol{\theta}) &:= (\mathbf{I}_{p^2} + \mathbf{K}_{pp})\{\mathbf{U}(\boldsymbol{\varphi})\mathbf{M}(\boldsymbol{\mu}) \otimes \mathbf{I}_p\}D\mathbf{U}(\boldsymbol{\varphi}), \\ D_{\boldsymbol{\mu}}\boldsymbol{\Sigma}(\boldsymbol{\theta}) &:= \{\mathbf{U}(\boldsymbol{\varphi}) \otimes \mathbf{U}(\boldsymbol{\varphi})\}\mathbb{D}_r, \\ D\boldsymbol{\Sigma}(\boldsymbol{\theta}) &:= [D_{\boldsymbol{\varphi}}\boldsymbol{\Sigma}(\boldsymbol{\theta}) \quad D_{\boldsymbol{\mu}}\boldsymbol{\Sigma}(\boldsymbol{\theta})], \end{aligned}$$

where $D\mathbf{U}(\boldsymbol{\varphi})$ is the Fréchet derivative of the Cayley parameterization defined by (2.3). When $\boldsymbol{\theta}$ takes value at the referential Euclidean vector $\boldsymbol{\theta}_0 = [\boldsymbol{\varphi}_0^T, \boldsymbol{\mu}_0^T]^T$ such that $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0) = \mathbf{U}(\boldsymbol{\varphi}_0)\mathbf{M}(\boldsymbol{\mu}_0)\mathbf{U}(\boldsymbol{\varphi}_0)^T$, we simply write $\mathbf{U}_0 = \mathbf{U}(\boldsymbol{\varphi}_0)$ and $\mathbf{M}_0 = \mathbf{M}(\boldsymbol{\mu}_0)$.

3. Main results.

3.1. Intrinsic perturbation theorems. We present our first main result in Theorem 3.1 below, which translates the perturbation of two matrices $\Sigma(\theta)$, $\Sigma(\theta_0)$ on the same manifold $\mathcal{S}(p, r)$ to the perturbation of the corresponding representing Euclidean vectors θ , θ_0 through a first-order Taylor expansion device.

THEOREM 3.1. *Under the setup and notations of Sections 2.2 and 2.4, if $\|\mathbf{A}_0\|_2 < 1$ and $\|\mathbf{A}\|_2 < 1$, then there exists a $p \times p$ matrix $\mathbf{R}(\theta, \theta_0)$ depending on θ, θ_0 , such that*

$$\begin{aligned} \Sigma(\theta) - \Sigma_0 &= 2(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}(\mathbf{X}_{\varphi} - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-T}\Sigma_0 \\ &\quad - 2\Sigma_0(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}(\mathbf{X}_{\varphi} - \mathbf{X}_{\varphi_0})(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-T} \\ &\quad + \mathbf{U}_0\{\mathbf{M}(\mu) - \mathbf{M}_0\}\mathbf{U}_0^T + \mathbf{R}(\theta, \theta_0), \end{aligned}$$

where the remainder $\mathbf{R}(\theta, \theta_0)$ satisfies $\|\mathbf{R}(\theta, \theta_0)\|_F \leq 16(1 + \|\mathbf{M}_0\|_2)\|\theta - \theta_0\|_2^2$.

Proof. First observe that the following matrix decomposition holds:

$$\begin{aligned} \Sigma(\theta) - \Sigma(\theta_0) &= \mathbf{U}_0(\mathbf{M} - \mathbf{M}_0)\mathbf{U}_0^T + \mathbf{U}_0\mathbf{M}_0\{\mathbf{U}(\varphi) - \mathbf{U}_0\}^T + \{\mathbf{U}(\varphi) - \mathbf{U}_0\}\mathbf{M}_0\mathbf{U}_0^T + \mathbf{R}_{\Sigma}(\theta, \theta_0), \end{aligned}$$

where the remainder

$$\begin{aligned} \mathbf{R}_{\Sigma}(\theta, \theta_0) &= \mathbf{U}(\varphi)(\mathbf{M} - \mathbf{M}_0)\{\mathbf{U}(\varphi) - \mathbf{U}_0\}^T + \{\mathbf{U}(\varphi) - \mathbf{U}_0\}\mathbf{M}_0\{\mathbf{U}(\varphi) - \mathbf{U}_0\}^T \\ &\quad + \{\mathbf{U}(\varphi) - \mathbf{U}_0\}(\mathbf{M} - \mathbf{M}_0)\mathbf{U}_0^T \end{aligned}$$

satisfies

$$\begin{aligned} \|\mathbf{R}_{\Sigma}(\theta, \theta_0)\|_F &\leq 2\|\mathbf{M} - \mathbf{M}_0\|_F\|\mathbf{U}(\varphi) - \mathbf{U}_0\|_F + \|\mathbf{M}_0\|_2\|\mathbf{U}(\varphi) - \mathbf{U}_0\|_F^2 \\ &\leq \|\mathbf{M} - \mathbf{M}_0\|_2^2 + \|\mathbf{U}(\varphi) - \mathbf{U}_0\|_F^2 + \|\mathbf{M}_0\|_2\|\mathbf{U}(\varphi) - \mathbf{U}_0\|_F^2 \\ &\leq 2\|\mu - \mu_0\|_2^2 + (1 + \|\mathbf{M}_0\|_2)\|\mathbf{U}(\varphi) - \mathbf{U}_0\|_F^2. \end{aligned}$$

By Theorem 2.1,

$$\|\mathbf{U}(\varphi) - \mathbf{U}_0\|_F \leq 2\sqrt{2}\|\varphi - \varphi_0\|_2$$

for all φ and φ_0 . Therefore,

$$\|\mathbf{R}_{\Sigma}(\theta, \theta_0)\|_F \leq 8(1 + \|\mathbf{M}_0\|_2)\|\theta - \theta_0\|_2^2.$$

Furthermore, using Theorem 2.1 again, we obtain

$$\text{vec}\{\mathbf{U}(\varphi) - \mathbf{U}_0\} = D\mathbf{U}(\varphi_0)(\varphi - \varphi_0) + \text{vec}\{\mathbf{R}_{\mathbf{U}}(\varphi, \varphi_0)\},$$

where $\|\mathbf{R}_{\mathbf{U}}(\varphi, \varphi_0)\|_F \leq 8\|\varphi - \varphi_0\|_2^2$ for all φ, φ_0 . In matrix form, we have

$$\mathbf{U}(\varphi) - \mathbf{U}_0 = 2(\mathbf{I}_p - \mathbf{X}_0)^{-1}(\mathbf{X}_{\varphi} - \mathbf{X}_0)(\mathbf{I}_p - \mathbf{X}_0)^{-1}\mathbf{I}_{p \times r} + \mathbf{R}_{\mathbf{U}}(\varphi, \varphi_0).$$

Hence we finally obtain

$$\begin{aligned} \text{vec}\{\Sigma(\theta) - \Sigma_0\} &= D_{\mu}\Sigma(\theta_0)(\mu - \mu_0) + (\mathbf{I}_{p^2} + \mathbf{K}_{pp})(\mathbf{U}_0\mathbf{M}_0 \otimes \mathbf{I}_p)\text{vec}\{\mathbf{U}(\varphi) - \mathbf{U}_0\} \\ &\quad + \text{vec}\{\mathbf{R}_{\Sigma}(\theta, \theta_0)\} \\ &= D\Sigma(\theta_0)(\theta - \theta_0) + (\mathbf{I}_p^2 + \mathbf{K}_{pp})(\mathbf{U}_0\mathbf{M}_0 \otimes \mathbf{I}_p)\text{vec}\{\mathbf{R}_{\mathbf{U}}(\varphi, \varphi_0)\} \\ &\quad + \text{vec}\{\mathbf{R}_{\Sigma}(\theta, \theta_0)\} \\ &= D\Sigma(\theta_0)(\theta - \theta_0) + \text{vec}\{\mathbf{R}(\theta, \theta_0)\}, \end{aligned}$$

where

$$\mathbf{R}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = (\mathbf{I}_p^2 + \mathbf{K}_{pp})(\mathbf{U}_0 \mathbf{M}_0 \otimes \mathbf{I}_p) \text{vec}\{\mathbf{R}_{\mathbf{U}}(\boldsymbol{\varphi}, \boldsymbol{\varphi}_0)\} + \text{vec}\{\mathbf{R}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0)\},$$

and the proof is completed by observing that

$$\|\mathbf{R}(\boldsymbol{\theta}, \boldsymbol{\theta}_0)\|_{\text{F}} \leq 2\|\mathbf{M}_0\|_2 \|\mathbf{R}_{\mathbf{U}}(\boldsymbol{\varphi}, \boldsymbol{\varphi}_0)\|_{\text{F}} + \|\mathbf{R}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0)\|_{\text{F}} \leq 16(1 + \|\mathbf{M}_0\|_2) \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2^2. \square$$

Theorem 3.1 immediately implies that, locally at $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, the Frobenius norm of the perturbation $\|\mathbf{E}\|_{\text{F}} = \|\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|_{\text{F}}$ between $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ on the same low-rank matrix manifold $\mathcal{S}(p, r)$ can be well controlled by the Euclidean distance between their representing Euclidean vector. Theorem 3.2 below, which is our second main result, asserts that the reverse statement is true: the perturbation of the representing Euclidean vectors can be well controlled by the perturbation of the original matrices locally at $\boldsymbol{\Sigma}_0$.

THEOREM 3.2. *Under the setup and notations of Sections 2.2 and 2.4, if $\|\mathbf{A}_0\|_2 < 1$, $\|\mathbf{A}\|_2 < 1$, \mathbf{M}_0 and \mathbf{M} are positive definite, and*

$$\|\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|_{\text{F}} \leq \frac{(1 - \|\mathbf{A}_0\|_2^2)^2 \lambda_r(\mathbf{M}_0)}{4\sqrt{2}(1 + \|\mathbf{A}_0\|_2^2)^2},$$

then

$$\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 \leq \left[1 + \frac{16\sqrt{2}\{1 + \lambda_1(\mathbf{M}_0)\}(1 + \|\mathbf{A}_0\|_2^2)}{\lambda_r(\mathbf{M}_0)(1 - \|\mathbf{A}_0\|_2^2)} \right] \|\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|_{\text{F}}.$$

Before proving Theorem 3.2, we introduce the following intermediate Lemma claiming that the perturbation of projection matrices can be controlled by the corresponding Euclidean representing vectors. Its proof is provided in Appendix.

LEMMA 3.3. *Under the setup of Section 2.2, if $\|\mathbf{A}\|_2, \|\mathbf{A}_0\|_2 < 1$, then*

$$\begin{aligned} \|\mathbf{U}(\boldsymbol{\varphi}) - \mathbf{U}(\boldsymbol{\varphi}_0)\|_{\text{F}} &\leq \frac{2\|\mathbf{U}(\boldsymbol{\varphi})\mathbf{U}(\boldsymbol{\varphi})^{\text{T}} - \mathbf{U}(\boldsymbol{\varphi}_0)\mathbf{U}(\boldsymbol{\varphi}_0)^{\text{T}}\|_{\text{F}}}{\min\{\lambda_r(\mathbf{Q}_{01}), \lambda_r(\mathbf{Q}_1)\}}, \\ \|\boldsymbol{\varphi} - \boldsymbol{\varphi}_0\|_2 &\leq \frac{4\|\mathbf{U}(\boldsymbol{\varphi})\mathbf{U}(\boldsymbol{\varphi})^{\text{T}} - \mathbf{U}(\boldsymbol{\varphi}_0)\mathbf{U}(\boldsymbol{\varphi}_0)^{\text{T}}\|_{\text{F}}}{\min\{\lambda_r(\mathbf{Q}_{01}), \lambda_r(\mathbf{Q}_1)\}}. \end{aligned}$$

Proof of Theorem 3.2. By Weyl's inequality, we have

$$|\lambda_r(\mathbf{Q}_1) - \lambda_r(\mathbf{Q}_{01})| = \frac{|\lambda_r^2(\mathbf{Q}_1) - \lambda_r^2(\mathbf{Q}_{01})|}{\lambda_r(\mathbf{Q}_1) + \lambda_r(\mathbf{Q}_{01})} \leq \frac{\|\mathbf{Q}_1^2 - \mathbf{Q}_{01}^2\|_{\text{F}}}{\lambda_r(\mathbf{Q}_{01})} \leq \frac{\|\mathbf{U}\mathbf{U}^{\text{T}} - \mathbf{U}_0\mathbf{U}_0^{\text{T}}\|_{\text{F}}}{\lambda_r(\mathbf{Q}_{01})}.$$

Note that $\lambda_r(\mathbf{Q}_{01}) = \lambda_r\{(\mathbf{I}_r - \mathbf{A}_0^{\text{T}}\mathbf{A}_0)(\mathbf{I}_r + \mathbf{A}_0^{\text{T}}\mathbf{A}_0)^{-1}\} = (1 - \|\mathbf{A}_0\|_2^2)/(1 + \|\mathbf{A}_0\|_2^2)$. By the Davis-Kahan theorem (see, e.g., Theorem 2 in [62]),

$$|\lambda_r(\mathbf{Q}_1) - \lambda_r(\mathbf{Q}_{01})| \leq \frac{\|\mathbf{U}\mathbf{U}^{\text{T}} - \mathbf{U}_0\mathbf{U}_0^{\text{T}}\|_{\text{F}}}{\lambda_r(\mathbf{Q}_{01})} \leq 2\sqrt{2} \left(\frac{1 + \|\mathbf{A}_0\|_2^2}{1 - \|\mathbf{A}_0\|_2^2} \right) \frac{\|\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|_{\text{F}}}{\lambda_r(\mathbf{M}_0)}.$$

Therefore,

$$\begin{aligned} \lambda_r(\mathbf{Q}_1) &\geq \frac{1 - \|\mathbf{A}_0\|_2^2}{1 + \|\mathbf{A}_0\|_2^2} - \frac{2\sqrt{2}(1 + \|\mathbf{A}_0\|_2^2)\|\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|_{\text{F}}}{(1 - \|\mathbf{A}_0\|_2^2)\lambda_r(\mathbf{M}_0)} \\ &\geq \frac{1 - \|\mathbf{A}_0\|_2^2}{2(1 + \|\mathbf{A}_0\|_2^2)} = \frac{\lambda_r(\mathbf{Q}_{01})}{2}. \end{aligned}$$

Hence, by Lemma 3.3 and the Davis-Kahan theorem, we have

$$\begin{aligned} \|\mathbf{U} - \mathbf{U}_0\|_F &\leq \frac{4\|\mathbf{U}\mathbf{U}^T - \mathbf{U}_0\mathbf{U}_0^T\|_F}{\lambda_r(\mathbf{Q}_{01})} \leq \frac{8\sqrt{2}\|\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|_F}{\lambda_r(\mathbf{Q}_{01})\lambda_r(\mathbf{M}_0)} \\ &= \frac{8\sqrt{2}(1 + \|\mathbf{A}_0\|_2^2)}{\lambda_r(\mathbf{M}_0)(1 - \|\mathbf{A}_0\|_2^2)} \|\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|_F, \\ \|\boldsymbol{\varphi} - \boldsymbol{\varphi}_0\|_2 &\leq \frac{8\|\mathbf{U}\mathbf{U}^T - \mathbf{U}_0\mathbf{U}_0^T\|_F}{\lambda_r(\mathbf{Q}_{01})} \leq \frac{16\sqrt{2}\|\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|_F}{\lambda_r(\mathbf{Q}_{01})\lambda_r(\mathbf{M}_0)} \\ &= \frac{16\sqrt{2}(1 + \|\mathbf{A}_0\|_2^2)}{\lambda_r(\mathbf{M}_0)(1 - \|\mathbf{A}_0\|_2^2)} \|\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|_F. \end{aligned}$$

For the matrix \mathbf{M} and the vector $\boldsymbol{\mu}$, we have,

$$\begin{aligned} \|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|_2 &\leq \|\mathbf{M} - \mathbf{M}_0\|_F = \|\mathbf{U}^T \boldsymbol{\Sigma}(\boldsymbol{\theta}) \mathbf{U} - \mathbf{U}_0^T \boldsymbol{\Sigma}(\boldsymbol{\theta}_0) \mathbf{U}_0\|_F \\ &\leq \|\mathbf{U}^T \{\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\} \mathbf{U}\|_F + \|\mathbf{U}^T \boldsymbol{\Sigma}_0(\mathbf{U} - \mathbf{U}_0)\|_F + \|(\mathbf{U} - \mathbf{U}_0)^T \boldsymbol{\Sigma}_0 \mathbf{U}_0\|_F \\ &\leq \|\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|_F + 2\lambda_1(\mathbf{M}_0) \|\mathbf{U} - \mathbf{U}_0\|_F \\ &\leq \left\{ 1 + \frac{16\sqrt{2}\lambda_1(\mathbf{M}_0)(1 + \|\mathbf{A}_0\|_2^2)}{\lambda_r(\mathbf{M}_0)(1 - \|\mathbf{A}_0\|_2^2)} \right\} \|\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|_F. \end{aligned}$$

Therefore, we conclude that

$$\begin{aligned} \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 &\leq \|\boldsymbol{\varphi} - \boldsymbol{\varphi}_0\|_2 + \|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|_2 \\ &\leq \left[1 + \frac{16\sqrt{2}\{1 + \lambda_1(\mathbf{M}_0)\}(1 + \|\mathbf{A}_0\|_2^2)^2}{\lambda_r(\mathbf{M}_0)(1 - \|\mathbf{A}_0\|_2^2)^2} \right] \|\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|_F. \quad \square \end{aligned}$$

When $\mathbf{M} = \mathbf{M}_0 = \mathbf{I}_r$, Theorem 3.2, together with the fact that $\|\mathbf{U}\mathbf{U}^T - \mathbf{V}\mathbf{V}^T\|_F = \sqrt{2}\|\sin \Theta(\mathbf{U}, \mathbf{V})\|_F$ for any $\mathbf{U}, \mathbf{V} \in \mathbb{O}(p, r)$, directly leads to the following corollary regarding intrinsic perturbation of subspaces.

COROLLARY 3.4. *Under the setup and notations of Section 2.2, if $\|\mathbf{A}_0\|_2 < 1$, $\|\mathbf{A}\|_2 < 1$, and*

$$\|\sin \Theta\{\mathbf{U}(\boldsymbol{\varphi}), \mathbf{U}(\boldsymbol{\varphi}_0)\}\|_F \leq \frac{(1 - \|\mathbf{A}_0\|_2^2)^2}{8(1 + \|\mathbf{A}_0\|_2^2)^2},$$

then

$$\|\boldsymbol{\varphi} - \boldsymbol{\varphi}_0\|_2 \leq \sqrt{2} \left\{ 1 + \frac{32\sqrt{2}(1 + \|\mathbf{A}_0\|_2^2)^2}{(1 - \|\mathbf{A}_0\|_2^2)^2} \right\} \|\sin \Theta\{\mathbf{U}(\boldsymbol{\varphi}), \mathbf{U}(\boldsymbol{\varphi}_0)\}\|_F.$$

On the other hand, the following reverse inequality always holds for all $\boldsymbol{\varphi}$ and $\boldsymbol{\varphi}_0$:

$$\|\sin \Theta\{\mathbf{U}(\boldsymbol{\varphi}), \mathbf{U}(\boldsymbol{\varphi}_0)\}\|_F \leq 4\|\boldsymbol{\varphi} - \boldsymbol{\varphi}_0\|_2.$$

Corollary 3.4 suggests that, locally around $\boldsymbol{\varphi}_0$, the Frobenius sine-theta distance between $\text{Span}\{\mathbf{U}(\boldsymbol{\varphi})\}$ and $\text{Span}\{\mathbf{U}(\boldsymbol{\varphi}_0)\}$ is equivalent to the Euclidean distance between their representing Euclidean vectors $\boldsymbol{\varphi}$ and $\boldsymbol{\varphi}_0$. Furthermore, by taking Theorem 2.1 into consideration, we conclude directly that $\|\sin \Theta\{\mathbf{U}(\boldsymbol{\varphi}), \mathbf{U}(\boldsymbol{\varphi}_0)\}\|_F$ is locally equivalent to $\|\mathbf{U}(\boldsymbol{\varphi}) - \mathbf{U}(\boldsymbol{\varphi}_0)\|_F$. This result is formally stated in the following Theorem for ease of reference.

THEOREM 3.5. Under the setup and notations of Section 2.2, if $\|\mathbf{A}_0\|_2 < 1$, $\|\mathbf{A}\|_2 < 1$, and

$$\|\sin \Theta\{\mathbf{U}(\varphi), \mathbf{U}(\varphi_0)\}\|_F \leq \frac{(1 - \|\mathbf{A}_0\|_2^2)^2}{8(1 + \|\mathbf{A}_0\|_2^2)^2},$$

then

$$\|\mathbf{U}(\varphi) - \mathbf{U}(\varphi_0)\|_F \leq 4 \left\{ 1 + \frac{32\sqrt{2}(1 + \|\mathbf{A}_0\|_2^2)^2}{(1 - \|\mathbf{A}_0\|_2^2)^2} \right\} \|\sin \Theta\{\mathbf{U}(\varphi), \mathbf{U}(\varphi_0)\}\|_F.$$

On the other hand, the following reverse inequality always holds for all φ and φ_0 :

$$\|\sin \Theta\{\mathbf{U}(\varphi), \mathbf{U}(\varphi_0)\}\|_F \leq \sqrt{2} \|\mathbf{U}(\varphi) - \mathbf{U}(\varphi_0)\|_F.$$

Given two Stiefel matrices \mathbf{V} and \mathbf{V}_0 in $\mathbb{O}(p, r)$, their Frobenius sine-theta distance is equivalent to $\|\mathbf{V} - \mathbf{V}_0 \mathbf{W}^*\|_F$, where \mathbf{W}^* solves the orthogonal Procrustes problem $\min_{\mathbf{W} \in \mathbb{O}(r)} \|\mathbf{V} - \mathbf{V}_0 \mathbf{W}\|_F$ and can be computed explicitly using \mathbf{V}_0 and \mathbf{V} . Formally, by Lemma 4.1 in [64] (also see Lemma 1 in [14]), we have

$$\|\sin \Theta(\mathbf{V}, \mathbf{V}_0)\|_F \leq \|\mathbf{V} - \mathbf{V}_0 \mathbf{W}^*\|_F \leq \sqrt{2} \|\sin \Theta(\mathbf{V}, \mathbf{V}_0)\|_F.$$

As pointed out in [14], it is sometimes more convenient to work with the explicit expression $\|\mathbf{V} - \mathbf{V}_0 \mathbf{W}^*\|_F$ based on the representing Stiefel matrices than to work with the original definition of the sine-theta distance between subspaces. Nevertheless, the orthogonal alignment matrix \mathbf{W}^* may still cause inconvenience for some theoretical analyses. In contrast, Theorem 3.5 loosely asserts that, in a small neighborhood of \mathbf{V}_0 (or equivalently, a small neighborhood of $\text{Span}(\mathbf{V}_0)$ with respect to the sine-theta metric), by finding suitable representing Stiefel matrices $\mathbf{U}(\varphi), \mathbf{U}(\varphi_0) \in \mathbb{O}_+(p, r)$ such that $\text{Span}\{\mathbf{U}(\varphi)\} = \text{Span}(\mathbf{V})$ and $\text{Span}\{\mathbf{U}(\varphi_0)\} = \text{Span}(\mathbf{V}_0)$, $\|\sin \Theta(\mathbf{V}, \mathbf{V}_0)\|_F$ is equivalent to $\|\mathbf{U}(\varphi) - \mathbf{U}(\varphi_0)\|_F$, circumventing the orthogonal alignment matrix \mathbf{W}^* and facilitating many technical analyses. Therefore, the metric $\|\mathbf{U}(\varphi) - \mathbf{U}(\varphi_0)\|_F$ provides an alignment-free and user-friendly local surrogate for the Frobenius sine-theta distance.

3.2. The regularity theorem. Theorem 3.6 below is the third main result of this work. It asserts that the map $\Sigma(\cdot) : \mathcal{D}(p, r) \rightarrow \mathcal{S}(p, r)$ is regular by showing that the Fréchet derivative $D\Sigma(\cdot)$ has full column rank. Furthermore, $\sigma_{\min}\{D\Sigma(\theta_0)\}$ can be lower bounded using $\sigma_r(\mathbf{M}_0)$ and $\|\mathbf{A}_0\|_2$.

THEOREM 3.6. Under the setup and notations of Sections 2.2 and 2.4,

$$\sigma_{\min}\{D_\varphi \Sigma(\theta_0)\} \geq \begin{cases} \frac{2\sqrt{2}\sigma_r(\mathbf{M}_0)(1 - \|\mathbf{A}_0\|_2^2)}{(1 + \|\mathbf{A}_0\|_2^2)^2}, & \text{if } r \geq 2, \\ \frac{2\sqrt{2}\sigma_r(\mathbf{M}_0)}{1 + \|\mathbf{A}_0\|_2^2}, & \text{if } r = 1. \end{cases}$$

Furthermore,

$$\|\{D\Sigma(\theta_0)^T D\Sigma(\theta_0)\}^{-1}\|_2 \leq \begin{cases} 1 + \frac{(1 + 64\|\mathbf{M}_0\|_2^2)(1 + \|\mathbf{A}_0\|_2^2)^4}{8\lambda_r^2(\mathbf{M}_0)(1 - \|\mathbf{A}_0\|_2^2)^2}, & \text{if } r \geq 2, \\ 1 + \frac{(1 + 64\|\mathbf{M}_0\|_2^2)(1 + \|\mathbf{A}_0\|_2^2)^2}{8\lambda_r^2(\mathbf{M}_0)}, & \text{if } r = 1. \end{cases}$$

The proof of Theorem 3.6 is involved and relies on the following two technical lemmas, the proofs of which are deferred to Appendix.

LEMMA 3.7. *Under the setup and notations in Sections 2.2 and 2.4,*

$$\begin{aligned} & \Sigma_0^2 \otimes \mathbf{I}_p - \Sigma_0 \otimes \Sigma_0 + \mathbf{I}_p \otimes \Sigma_0^2 \\ & \geq \lambda_r^2(\mathbf{M}_0) \{ \mathbf{U}_0 \mathbf{U}_0^T \otimes (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) + (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) \otimes \mathbf{U}_0 \mathbf{U}_0^T \}, \\ & \Sigma_0^2 \otimes (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) + (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) \otimes \Sigma_0^2 \\ & \geq \lambda_r^2(\mathbf{M}_0) \{ \mathbf{U}_0 \mathbf{U}_0^T \otimes (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) + (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) \otimes \mathbf{U}_0 \mathbf{U}_0^T \}. \end{aligned}$$

LEMMA 3.8. *Let $\mathbf{A}_0 \in \mathbb{R}^{(p-r) \times r}$ with $\|\mathbf{A}_0\|_2 < 1$, and define*

$$(3.2) \quad \begin{aligned} \mathbf{C}_{11} &= (\mathbf{I}_r + \mathbf{A}_0^T \mathbf{A}_0)^{-1}, & \mathbf{C}_{12} &= -(\mathbf{I}_r + \mathbf{A}_0^T \mathbf{A}_0)^{-1} \mathbf{A}_0^T, \\ \mathbf{C}_{21} &= \mathbf{A}_0 (\mathbf{I}_r + \mathbf{A}_0^T \mathbf{A}_0)^{-1}, & \mathbf{C}_{22} &= \mathbf{I}_{p-r} - \mathbf{A}_0 (\mathbf{I}_r + \mathbf{A}_0^T \mathbf{A}_0)^{-1} \mathbf{A}_0^T. \end{aligned}$$

Let

$$\mathbf{C}_0 = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}, \quad \text{and} \quad \mathbf{U}_0 = \mathbf{C}_0^{-T} \mathbf{C}_0 \mathbf{I}_{p \times r}.$$

(i) *For any vector $\mathbf{A} \in \mathbb{R}^{(p-r) \times r}$ and $\boldsymbol{\varphi} := \text{vec}(\mathbf{A})$,*

$$\begin{aligned} & \text{vec}(\mathbf{X}_{\boldsymbol{\varphi}})^T (\mathbf{C}_0^T \otimes \mathbf{C}_0^T) \{ \mathbf{U}_0 \mathbf{U}_0 \otimes (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) \} (\mathbf{C}_0 \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_{\boldsymbol{\varphi}}) \\ & = \text{vec}(\mathbf{X}_{\boldsymbol{\varphi}})^T (\mathbf{C}_0^T \otimes \mathbf{C}_0^T) \{ (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) \otimes \mathbf{U}_0 \mathbf{U}_0^T \} (\mathbf{C}_0 \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_{\boldsymbol{\varphi}}) \\ & = \|\mathbf{C}_{22}^T \mathbf{A} \mathbf{C}_{11} - \mathbf{C}_{12}^T \mathbf{A}^T \mathbf{C}_{21}\|_F^2, \end{aligned}$$

where

$$\mathbf{X}_{\boldsymbol{\varphi}} := \begin{bmatrix} r \times r & -\mathbf{A}^T \\ \mathbf{A} & (p-r) \times (p-r) \end{bmatrix}.$$

(ii) $\mathbf{C}_{11} \otimes \mathbf{C}_{22} - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T) \mathbf{K}_{(p-r)r}$ *has the following lower bound in spectra:*

$$\sigma_{\min} \{ \mathbf{C}_{11} \otimes \mathbf{C}_{22} - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T) \mathbf{K}_{(p-r)r} \} \geq \begin{cases} \frac{1 - \|\mathbf{A}_0\|_2^2}{(1 + \|\mathbf{A}_0\|_2^2)^2}, & \text{if } r > 1, \\ \frac{1}{1 + \|\mathbf{A}_0\|_2^2}, & \text{if } r = 1. \end{cases}$$

Proof of first assertion of Theorem 3.6. Let \mathbf{A} be any $(p-r) \times r$ matrix and $\boldsymbol{\varphi} := \text{vec}(\mathbf{A})$. Denote

$$\mathbf{X}_{\boldsymbol{\varphi}} = \begin{bmatrix} r \times r & -\mathbf{A}^T \\ \mathbf{A} & (p-r) \times (p-r) \end{bmatrix}.$$

Write $\mathbf{X}_0 = \mathbf{X}_{\boldsymbol{\varphi}_0}$ and $\mathbf{C}_0 = (\mathbf{I}_p - \mathbf{X}_0)^{-1}$. Let \mathbf{A}_0 be the corresponding $(p-r) \times r$ matrix such that $\boldsymbol{\varphi}_0 = \text{vec}(\mathbf{A}_0)$. Then for any $\boldsymbol{\varphi} = \text{vec}(\mathbf{A}) \in \mathbb{R}^d$, we have $\boldsymbol{\Gamma}_{\boldsymbol{\varphi}} \boldsymbol{\varphi} = \text{vec}(\mathbf{X}_{\boldsymbol{\varphi}})$ by the definition of $\boldsymbol{\Gamma}_{\boldsymbol{\varphi}}$. Therefore,

$$\begin{aligned} & (\mathbf{U}_0 \mathbf{M}_0 \otimes \mathbf{I}_r) D\mathbf{U}(\boldsymbol{\varphi}_0) \boldsymbol{\varphi} = 2(\mathbf{U}_0 \mathbf{M}_0 \otimes \mathbf{I}_r) [\mathbf{I}_{p \times r}^T (\mathbf{I}_p - \mathbf{X}_0)^{-T} \otimes (\mathbf{I} - \mathbf{X}_0)^{-1}] \text{vec}(\mathbf{X}_{\boldsymbol{\varphi}}) \\ & = 2[\mathbf{U}_0 \mathbf{M}_0 \mathbf{I}_{p \times r}^T (\mathbf{I}_p - \mathbf{X}_0)^{-T} \otimes (\mathbf{I} - \mathbf{X}_0)^{-1}] \text{vec}(\mathbf{X}_{\boldsymbol{\varphi}}) \\ & = 2 \text{vec} \{ (\mathbf{I}_p - \mathbf{X}_0)^{-1} \mathbf{X}_{\boldsymbol{\varphi}} (\mathbf{I}_p - \mathbf{X}_0)^{-1} \mathbf{I}_{p \times r} \mathbf{M}_0 \mathbf{U}_0^T \} \\ & = 2 \text{vec} \{ (\mathbf{I}_p - \mathbf{X}_0)^{-1} \mathbf{X}_{\boldsymbol{\varphi}} (\mathbf{I}_p - \mathbf{X}_0)^{-T} \mathbf{U}_0 \mathbf{M}_0 \mathbf{U}_0^T \}. \end{aligned}$$

By definition of the commutation matrix \mathbf{K}_{pp} , $\mathbf{K}_{pp}\text{vec}(\mathbf{M}) = \text{vec}(\mathbf{M})$ for any symmetric $\mathbf{M} \in \mathbb{R}^{p \times p}$ and $\mathbf{K}_{pp}\text{vec}(\mathbf{M}) = -\text{vec}(\mathbf{M})$ for any skew-symmetric $\mathbf{M} \in \mathbb{R}^{p \times p}$. Hence, for any $p \times p$ matrix \mathbf{M} , we have, $(\mathbf{I}_{p^2} + \mathbf{K}_{pp})\text{vec}(\mathbf{M}) = \text{vec}(\mathbf{M} + \mathbf{M}^T)$, and hence,

$$\begin{aligned} D_{\varphi}\Sigma(\theta_0)\varphi &= (\mathbf{I}_{p^2} + \mathbf{K}_{pp})(\mathbf{U}_0\mathbf{M}_0 \otimes \mathbf{I}_r)D\mathbf{U}(\varphi_0)\varphi \\ &= 2(\mathbf{I}_{p^2} + \mathbf{K}_{pp})\text{vec}\{(\mathbf{I}_p - \mathbf{X}_0)^{-1}\mathbf{X}_{\varphi}(\mathbf{I}_p - \mathbf{X}_0)^{-T}\mathbf{U}_0\mathbf{M}_0\mathbf{U}_0^T\} \\ &= 2\text{vec}(\mathbf{C}_0\mathbf{X}_{\varphi}\mathbf{C}_0^T\Sigma_0 - \Sigma_0\mathbf{C}_0\mathbf{X}_{\varphi}\mathbf{C}_0^T) \\ &= 2(\Sigma_0\mathbf{C}_0 \otimes \mathbf{C}_0 - \mathbf{C}_0 \otimes \Sigma_0\mathbf{C}_0)\text{vec}(\mathbf{X}_{\varphi}). \end{aligned}$$

Let $\mathbf{U}_{0\perp}$ be the orthogonal complement of \mathbf{U}_0 such that $[\mathbf{U}_0, \mathbf{U}_{0\perp}] \in \mathbb{O}(p)$. By Lemma 3.7,

$$\begin{aligned} \|D_{\varphi}\Sigma(\theta_0)\varphi\|_2^2 &= 4\text{vec}(\mathbf{X}_{\varphi})^T(\Sigma_0\mathbf{C}_0 \otimes \mathbf{C}_0 - \mathbf{C}_0 \otimes \Sigma_0\mathbf{C}_0)^T(\Sigma_0\mathbf{C}_0 \otimes \mathbf{C}_0 - \mathbf{C}_0 \otimes \Sigma_0\mathbf{C}_0)\text{vec}(\mathbf{X}_{\varphi}) \\ &= 4\text{vec}(\mathbf{X}_{\varphi})^T(\mathbf{C}_0^T\Sigma_0^2\mathbf{C}_0 \otimes \mathbf{C}_0^T\mathbf{C}_0 - 2\mathbf{C}_0^T\Sigma_0\mathbf{C}_0 \otimes \mathbf{C}_0^T\Sigma_0\mathbf{C}_0 + \mathbf{C}_0^T\mathbf{C}_0 \otimes \mathbf{C}_0^T\Sigma_0^2\mathbf{C}_0) \\ &\quad \times \text{vec}(\mathbf{X}_{\varphi}) \\ &= 4\text{vec}(\mathbf{X}_{\varphi})^T\{(\mathbf{C}_0^T \otimes \mathbf{C}_0^T)(\Sigma_0^2 \otimes \mathbf{I}_p - 2\Sigma_0 \otimes \Sigma_0 + \mathbf{I}_p \otimes \Sigma_0^2)(\mathbf{C}_0 \otimes \mathbf{C}_0)\}\text{vec}(\mathbf{X}_{\varphi}) \\ &\geq 4\lambda_r^2(\mathbf{M}_0)\text{vec}(\mathbf{X}_{\varphi})^T(\mathbf{C}_0^T \otimes \mathbf{C}_0^T)(\mathbf{U}_0\mathbf{U}_0^T \otimes \mathbf{U}_{0\perp}\mathbf{U}_{0\perp}^T + \mathbf{U}_{0\perp}\mathbf{U}_{0\perp}^T \otimes \mathbf{U}_0\mathbf{U}_0^T) \\ &\quad \times (\mathbf{C}_0 \otimes \mathbf{C}_0)\text{vec}(\mathbf{X}_{\varphi}) \\ &= 4\lambda_r^2\text{vec}(\mathbf{X}_{\varphi})^T(\mathbf{C}_0^T \otimes \mathbf{C}_0^T)\{(\mathbf{I}_p - \mathbf{U}_0\mathbf{U}_0^T) \otimes \mathbf{U}_0\mathbf{U}_0^T\}(\mathbf{C}_0 \otimes \mathbf{C}_0)\text{vec}(\mathbf{X}_{\varphi}) \\ &\quad + 4\lambda_r^2\text{vec}(\mathbf{X}_{\varphi})^T(\mathbf{C}_0^T \otimes \mathbf{C}_0^T)\{\mathbf{U}_0\mathbf{U}_0^T \otimes (\mathbf{I}_p - \mathbf{U}_0\mathbf{U}_0^T)\}(\mathbf{C}_0 \otimes \mathbf{C}_0)\text{vec}(\mathbf{X}_{\varphi}). \end{aligned}$$

Write \mathbf{C}_0 in the block form

$$\mathbf{C}_0 = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \begin{bmatrix} (\mathbf{I}_r + \mathbf{A}_0^T\mathbf{A}_0)^{-1} & -(\mathbf{I}_r + \mathbf{A}_0^T\mathbf{A}_0)^{-1}\mathbf{A}_0^T \\ \mathbf{A}_0(\mathbf{I}_r + \mathbf{A}_0^T\mathbf{A}_0)^{-1} & \mathbf{I}_{p-r} - \mathbf{A}_0(\mathbf{I}_r + \mathbf{A}_0^T\mathbf{A}_0)^{-1}\mathbf{A}_0^T \end{bmatrix}$$

according to Appendix C in [34]. Invoking Lemma 3.8 (i), we further obtain the following lower bound for $\|D_{\varphi}\Sigma(\theta_0)\varphi\|_2^2$:

$$\begin{aligned} \|D_{\varphi}\Sigma(\theta_0)\varphi\|_2^2 &\geq 8\lambda_r^2(\mathbf{M}_0)\|\mathbf{C}_{22}^T\mathbf{A}\mathbf{C}_{11} - \mathbf{C}_{12}^T\mathbf{A}^T\mathbf{C}_{21}\|_F^2 \\ &= 8\lambda_r^2(\mathbf{M}_0)\|(\mathbf{C}_{11}^T \otimes \mathbf{C}_{22}^T)\text{vec}(\mathbf{A}) - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T)\text{vec}(\mathbf{A}^T)\|_2^2 \\ &= 8\lambda_r^2(\mathbf{M}_0)\|(\mathbf{C}_{11}^T \otimes \mathbf{C}_{22}^T)\text{vec}(\mathbf{A}) - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T)\mathbf{K}_{(p-r)r}\text{vec}(\mathbf{A})\|_2^2 \\ &\geq 8\lambda_r^2(\mathbf{M}_0)\sigma_{\min}^2\{(\mathbf{C}_{11} \otimes \mathbf{C}_{22}) - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T)\mathbf{K}_{(p-r)r}\}\|\varphi\|_2^2, \end{aligned}$$

which immediately implies that

$$\sigma_{\min}\{D_{\varphi}\Sigma(\theta_0)\} \geq 2\sqrt{2}\sigma_r(\mathbf{M}_0)\sigma_{\min}\{(\mathbf{C}_{11} \otimes \mathbf{C}_{22}) - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T)\mathbf{K}_{(p-r)r}\}.$$

By Lemma 3.8 (ii), we finally obtain

$$\sigma_{\min}\{D_{\varphi}\Sigma(\theta_0)\} \geq \begin{cases} \frac{2\sqrt{2}\sigma_r(\mathbf{M}_0)(1 - \|\mathbf{A}_0\|_2^2)}{(1 + \|\mathbf{A}_0\|_2^2)^2}, & \text{if } r > 1, \\ \frac{2\sqrt{2}\sigma_r(\mathbf{M}_0)}{1 + \|\mathbf{A}_0\|_2^2}, & \text{if } r = 1. \end{cases}$$

completing the proof of the first assertion regarding $\sigma_{\min}\{D_{\varphi}\Sigma(\theta_0)\}$. \square

Proof of second assertion of Theorem 3.6. Write

$$D\Sigma(\theta_0)^T D\Sigma(\theta_0) = \begin{bmatrix} \mathbf{J}_1 & \mathbf{J}_2 \\ \mathbf{J}_2^T & \mathbf{J}_3 \end{bmatrix} := \begin{bmatrix} D_\varphi \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) & D_\varphi \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \\ D_\mu \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) & D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \end{bmatrix}.$$

Note that the Schur complement of the block $D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0)$ of the entire matrix $D\Sigma(\theta_0)^T D\Sigma(\theta_0)$ is given by

$$\begin{aligned} \mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T &= D_\varphi \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) \\ &\quad - D_\varphi \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \{D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0)\}^{-1} D_\mu \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0), \end{aligned}$$

We now assume that $\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T$ is invertible. Then by the block matrix inversion formula,

$$\begin{aligned} &\{D\Sigma(\theta_0)^T D\Sigma(\theta_0)\}^{-1} \\ &= \begin{bmatrix} (\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1} & -(\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1} \mathbf{J}_2 \mathbf{J}_3^{-1} \\ -\mathbf{J}_3^{-1} \mathbf{J}_2^T (\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1} & \mathbf{J}_3^{-1} + \mathbf{J}_3^{-1} \mathbf{J}_2^T (\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1} \mathbf{J}_2 \mathbf{J}_3^{-1} \end{bmatrix}, \end{aligned}$$

By construction,

$$\|\mathbf{J}_3^{-1}\|_2 = \|\{D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0)\}^{-1}\|_2 = \|\{\mathbb{D}_r^T(\mathbf{I}_r \otimes \mathbf{I}_r) \mathbb{D}_r\}^{-1}\| \leq 1,$$

where the last inequality is due to Theorem 4.4 in [42]. In addition,

$$\begin{aligned} \|\mathbf{J}_2\|_2 &\leq \|D\mathbf{U}(\varphi_0)^T (\mathbf{M}_0 \mathbf{U}_0^T \otimes \mathbf{I}_p) (\mathbf{I}_{p^2} + \mathbf{K}_{pp}) (\mathbf{U}_0 \otimes \mathbf{U}_0) \mathbb{D}_r\| \\ &\leq 2 \|D\mathbf{U}(\varphi_0)\|_2 \|\mathbf{M}_0\|_2 \|\mathbb{D}_r\|_2 \\ &\leq 8 \|\mathbf{M}_0\|_2 \|\mathbf{I}_{p \times r} (\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\|_2 \|(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\|_2 \leq 8 \|\mathbf{M}_0\|_2. \end{aligned}$$

Thus, by Lemma 3.4 of [10], we see that

$$\begin{aligned} \|\{D\Sigma(\theta_0)^T D\Sigma(\theta_0)\}^{-1}\|_2 &\leq \|(\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1}\|_2 \\ &\quad + \|\mathbf{J}_3^{-1} + \mathbf{J}_3^{-1} \mathbf{J}_2^T (\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1} \mathbf{J}_2 \mathbf{J}_3^{-1}\|_2 \\ &\leq \|(\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1}\|_2 \\ &\quad + \|\mathbf{J}_3^{-1}\|_2 + \|\mathbf{J}_3^{-1}\|_2^2 \|\mathbf{J}_2\|_2^2 \|(\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1}\|_2 \\ &\leq 1 + (1 + 64 \|\mathbf{M}_0\|_2^2) \|(\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1}\|_2. \end{aligned}$$

Therefore, it is sufficient to provide a lower bound for the smallest eigenvalue of $\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T$. For any φ , we follow the computation above and write

$$\begin{aligned} &\varphi^T (\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T) \varphi \\ &= \varphi^T D_\varphi \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) \varphi \\ &\quad - \varphi^T D_\varphi \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \{D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0)\}^{-1} D_\mu \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) \varphi \\ &= 4 \text{vec}(\mathbf{X}_\varphi)^T \{(\mathbf{C}_0^T \otimes \mathbf{C}_0^T)(\Sigma_0^2 \otimes \mathbf{I}_p - 2\Sigma_0 \otimes \Sigma_0 + \mathbf{I}_p \otimes \Sigma_0^2)(\mathbf{C}_0 \otimes \mathbf{C}_0)\} \text{vec}(\mathbf{X}_\varphi) \\ &\quad - 4 \text{vec}(\mathbf{X}_\varphi)(\mathbf{C}_0^T \otimes \mathbf{C}_0^T)(\Sigma_0 \otimes \mathbf{I}_p - \mathbf{I}_p \otimes \Sigma) \\ &\quad \times D_\mu \Sigma(\theta_0) \{D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0)\}^{-1} D_\mu \Sigma(\theta_0) \\ &\quad \times (\Sigma_0 \otimes \mathbf{I}_p - \mathbf{I}_p \otimes \Sigma)(\mathbf{C}_0 \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_\varphi). \end{aligned}$$

Note that

$$\begin{aligned} D_{\mu}\Sigma(\theta_0)\{D_{\mu}\Sigma(\theta_0)^T D_{\mu}\Sigma(\theta_0)\}^{-1} D_{\mu}\Sigma(\theta_0) \\ = (\mathbf{U}_0 \otimes \mathbf{U}_0) \mathbb{D}_r \{\mathbb{D}_r^T (\mathbf{I}_r \otimes \mathbf{I}_r) \mathbb{D}_r\}^{-1} \mathbb{D}_r^T (\mathbf{U}_0 \otimes \mathbf{U}_0)^T \end{aligned}$$

is a projection matrix. Since $\mathbb{D}_r \in \mathbb{R}^{r^2 \times r(r+1)/2}$ has full column rank and $r^2 \geq r(r+1)/2$, then

$$\mathbb{D}_r \{\mathbb{D}_r^T (\mathbf{I}_r \otimes \mathbf{I}_r) \mathbb{D}_r\}^{-1} \mathbb{D}_r \preceq \mathbf{I}_{r^2},$$

and hence,

$$D_{\mu}\Sigma(\theta_0)\{D_{\mu}\Sigma(\theta_0)^T D_{\mu}\Sigma(\theta_0)\}^{-1} D_{\mu}\Sigma(\theta_0) \preceq (\mathbf{U}_0 \otimes \mathbf{U}_0)(\mathbf{U}_0 \otimes \mathbf{U}_0)^T.$$

Therefore, we invoke Lemmas 3.7 and 3.8 and proceed to compute

$$\begin{aligned} & \varphi^T (\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T) \varphi \\ & \geq 4 \text{vec}(\mathbf{X}_{\varphi})^T \{(\mathbf{C}_0^T \otimes \mathbf{C}_0^T)(\Sigma_0^2 \otimes \mathbf{I}_p - 2\Sigma_0 \otimes \Sigma_0 + \mathbf{I}_p \otimes \Sigma_0^2)(\mathbf{C}_0 \otimes \mathbf{C}_0)\} \text{vec}(\mathbf{X}_{\varphi}) \\ & \quad - 4 \text{vec}(\mathbf{X}_{\varphi})(\mathbf{C}_0^T \otimes \mathbf{C}_0^T)(\Sigma_0 \otimes \mathbf{I}_p - \mathbf{I}_p \otimes \Sigma_0)(\mathbf{U}_0 \otimes \mathbf{U}_0)(\mathbf{U}_0 \otimes \mathbf{U}_0)^T \\ & \quad \times (\Sigma_0 \otimes \mathbf{I}_p - \mathbf{I}_p \otimes \Sigma_0) \\ & \quad \times (\mathbf{C}_0 \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_{\varphi}) \\ & = 4 \text{vec}(\mathbf{X}_{\varphi})^T \{(\mathbf{C}_0^T \otimes \mathbf{C}_0^T)(\Sigma_0^2 \otimes \mathbf{I}_p - 2\Sigma_0 \otimes \Sigma_0 + \mathbf{I}_p \otimes \Sigma_0^2)(\mathbf{C}_0 \otimes \mathbf{C}_0)\} \text{vec}(\mathbf{X}_{\varphi}) \\ & \quad - 4 \text{vec}(\mathbf{X}_{\varphi})(\mathbf{C}_0^T \otimes \mathbf{C}_0^T)(\Sigma_0^2 \otimes \mathbf{U}_0 \mathbf{U}_0^T - 2\Sigma_0 \otimes \Sigma_0 + \mathbf{U}_0 \mathbf{U}_0^T \otimes \Sigma_0^2) \\ & \quad \times (\mathbf{C}_0 \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_{\varphi}) \\ & = 4 \text{vec}(\mathbf{X}_{\varphi})^T (\mathbf{C}_0^T \otimes \mathbf{C}_0^T) \{ \Sigma_0^2 \otimes (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) + (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) \otimes \Sigma_0^2 \} \\ & \quad \times (\mathbf{C}_0 \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_{\varphi}) \\ & \geq 4 \lambda_r^2(\mathbf{M}_0) \text{vec}(\mathbf{X}_{\varphi})^T (\mathbf{C}_0^T \otimes \mathbf{C}_0^T) (\mathbf{U}_0 \mathbf{U}_0^T \otimes \mathbf{U}_{0\perp} \mathbf{U}_{0\perp}^T + \mathbf{U}_{0\perp} \mathbf{U}_{0\perp}^T \otimes \mathbf{U}_0 \mathbf{U}_0^T) \\ & \quad \times (\mathbf{C}_0 \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_{\varphi}) \\ & = 8 \lambda_r^2(\mathbf{M}_0) \| \mathbf{C}_{22}^T \mathbf{A} \mathbf{C}_{11} - \mathbf{C}_{12}^T \mathbf{A}^T \mathbf{C}_{21} \|_F^2 \\ & \geq 8 \lambda_r^2(\mathbf{M}_0) \sigma_{\min}^2 \{ (\mathbf{C}_{11} \otimes \mathbf{C}_{22}) - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T) \mathbf{K}_{(p-r)r} \} \|\varphi\|_2^2. \end{aligned}$$

It follows from Lemma 3.8 (ii) that

$$\lambda_{\min}(\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T) \geq \begin{cases} \frac{8 \lambda_r^2(\mathbf{M}_0)(1 - \|\mathbf{A}_0\|_2^2)}{(1 + \|\mathbf{A}_0\|_2^2)^2}, & \text{if } r > 1, \\ \frac{8 \lambda_r^2(\mathbf{M}_0)}{1 + \|\mathbf{A}_0\|_2^2}, & \text{if } r = 1. \end{cases}$$

Therefore, using (3.3),

$$\| \{D\Sigma(\theta_0)^T D\Sigma(\theta_0)\}^{-1} \|_2 \leq \begin{cases} 1 + \frac{(1 + 64\|\mathbf{M}_0\|_2^2)(1 + \|\mathbf{A}_0\|_2^2)^2}{8 \lambda_r^2(\mathbf{M}_0)(1 - \|\mathbf{A}_0\|_2^2)}, & \text{if } r > 1, \\ 1 + \frac{(1 + 64\|\mathbf{M}_0\|_2^2)(1 + \|\mathbf{A}_0\|_2^2)}{8 \lambda_r^2(\mathbf{M}_0)}, & \text{if } r = 1, \end{cases}$$

and the proof is thus completed. \square

We now apply the regularity theorem to establish the following important result: The class of all $p \times p$ symmetric matrices with ranks no greater than r , denoted by $\mathcal{M}(p, r)$, can be equipped with a differentiable manifold structure. This result is formally summarized in Proposition 3.9 below. Note that unlike the matrix class $\mathcal{S}(p, r)$, the first r rows of any $\Sigma \in \mathcal{M}(p, r)$ is not necessarily linearly independent.

PROPOSITION 3.9. *Let $\mathcal{M}(p, r)$ be the class of all $p \times p$ symmetric matrices with rank no greater than r equipped with the Euclidean topology. Then $\mathcal{M}(p, r)$ is a differentiable manifold of dimension $d = r(r + 1)/2 + (p - r)r$, with the following differentiable structure: Let $\mathcal{P}(p)$ be the collection of all $p \times p$ permutation matrices and for each $\mathbf{P} \in \mathcal{P}(p)$, define*

$$\begin{aligned} \mathbf{x}_{\mathbf{P}} : \mathcal{D}(p, r) &\rightarrow \mathcal{M}(p, r), \\ \boldsymbol{\theta} &\mapsto \mathbf{P}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{P}^T. \end{aligned}$$

Then $\{(\mathcal{D}(p, r), \mathbf{x}_{\mathbf{P}})\}_{\mathbf{P} \in \mathcal{P}(p)}$ forms a differentiable structure on $\mathcal{M}(p, r)$.

The proof of Proposition 3.9 is routine and deferred to Appendix. This result also suggests the possibility of additional geometric structures on the low-rank matrix manifold $\mathcal{M}(p, r)$ based on the atlas, including the tangent spaces, the Riemannian metric, connections, and geodesics.

4. Applications.

4.1. Riemannian structure of Grassmannian. The Grassmannian $\mathcal{G}(p, r)$, i.e., the collection of all r -dimensional subspaces in \mathbb{R}^p equipped with a differentiable structure, forms a smooth manifold (see, for example, [31, 58]). In [3], the authors established the rich Riemannian structures of $\mathcal{G}(p, r)$ using abstract geometric tools, including the horizontal lift of tangent vectors and Lie brackets, and obtained the Riemannian connection and geodesics. In [8] and [26], the authors derived the Riemannian metric and the geodesics using extrinsic matrix representation of tangent vectors and their horizontal lifts defined through the quotient geometry. Below, we show that the technical tools developed in Section 3 can also be applied to establish the Riemannian structure of $\mathcal{G}(p, r)$ equivalent to that obtained in [3, 8, 26].

Every element $\mathbb{S} \in \mathcal{G}(p, r)$ can be uniquely identified by a projection matrix $\mathbf{U}\mathbf{U}^T$ where $\mathbf{U} \in \mathbb{O}(p, r)$ such that $\mathbb{S} = \text{Span}(\mathbf{U})$. The topology on $\mathcal{G}(p, r)$ can be taken as the topology induced by the metric $\rho(\mathbf{U}_1\mathbf{U}_1^T, \mathbf{U}_2\mathbf{U}_2^T) = \|\mathbf{U}_1\mathbf{U}_1^T - \mathbf{U}_2\mathbf{U}_2^T\|_F$ for any $\mathbf{U}_1, \mathbf{U}_2 \in \mathbb{O}(p, r)$. Without loss of generality, we do not distinguish between the Grassmannian $\mathcal{G}(p, r)$ and the collection of all rank- r $p \times p$ projection matrices. A similar treatment was also adopted in [8]. We first establish the differentiable structure of $\mathcal{G}(p, r)$ in Proposition 4.1 using the technical tools developed in Section 2 and Section 3. The proof is routine and deferred to Appendix.

PROPOSITION 4.1. *Let $\mathcal{P}(p)$ be the collection of all $p \times p$ permutation matrices and $\mathcal{A} = \{\text{vec}(\mathbf{A}) : \mathbf{A} \in \mathbb{R}^{(p-r) \times r}, \|\mathbf{A}\|_2 < 1\}$. For each $\mathbf{P} \in \mathcal{P}(p)$, define*

$$(4.1) \quad \begin{aligned} \mathbf{y}_{\mathbf{P}} : \mathcal{A} &\rightarrow \mathcal{G}(p, r), \\ \boldsymbol{\varphi} = \text{vec}(\mathbf{A}) &\mapsto \mathbf{P}\mathbf{U}(\boldsymbol{\varphi})\mathbf{U}(\boldsymbol{\varphi})^T\mathbf{P}^T, \end{aligned}$$

where $\mathbf{U}(\cdot)$ is the Cayley parameterization given by (2.1). Then $\{(\mathcal{A}, \mathbf{y}_{\mathbf{P}})\}_{\mathbf{P} \in \mathcal{P}(p)}$ forms a differentiable structure on $\mathcal{G}(p, r)$.

We next construct the Riemannian metric on $\mathcal{G}(p, r)$ using the chart (4.1). In [8],

the authors showed that any vector $\Xi \in T_{\Pi}\mathcal{G}(p, r)$ has the form

$$\Xi = \mathbf{Q} \begin{bmatrix} r \times r & \mathbf{Y}^T \\ \mathbf{Y} & (p-r) \times (p-r) \end{bmatrix} \mathbf{Q}^T,$$

where $\Pi = \mathbf{Q}\mathbf{I}_{p \times r}\mathbf{I}_{p \times r}^T\mathbf{Q}^T$, $\mathbf{Y} \in \mathbb{R}^{(p-r) \times r}$, and for any $\Xi_i \in T_{\Pi}\mathcal{G}(p, r)$ with

$$\Xi_i = \mathbf{Q} \begin{bmatrix} r \times r & \mathbf{Y}_i^T \\ \mathbf{Y}_i & (p-r) \times (p-r) \end{bmatrix} \mathbf{Q}^T, \quad i = 1, 2,$$

the canonical metric is defined as $\langle \Xi_1, \Xi_2 \rangle_{\Pi} = \text{tr}(\mathbf{Y}_1^T \mathbf{Y}_2)$ (see, for example, [3, 8, 26]). By Lemma 3.8, for each $i = 1, 2$, there exists a unique matrix $\mathbf{B}_i \in \mathbb{R}^{(p-r) \times r}$ such that $\mathbf{Y}_i = \mathbf{C}_{22}\mathbf{B}_i\mathbf{C}_{11} - \mathbf{C}_{12}^T\mathbf{B}_i^T\mathbf{C}_{21}$, where $\mathbf{C}_{11}, \mathbf{C}_{12}, \mathbf{C}_{21}, \mathbf{C}_{22}$ are given by (3.2). Below, we take advantage of such a relation between \mathbf{B}_i and \mathbf{Y}_i to construct the Riemannian metric on $\mathcal{G}(p, r)$ using the Cayley parameterization.

PROPOSITION 4.2. *Let $\Pi \in \mathcal{G}(p, r)$ and suppose $(\mathcal{A}, \mathbf{y}_{\mathbf{P}})$ is a parameterization of $\mathcal{G}(p, r)$ at Π given by (4.1) for some $p \times p$ permutation matrix \mathbf{P} , such that $\mathbf{y}_{\mathbf{P}}(\text{vec}(\mathbf{A})) = \Pi$, where $\text{vec}(\mathbf{A}) \in \mathcal{A}$. Let $\Xi_i \in T_{\Pi}\mathcal{G}(p, r)$, $i = 1, 2$, have representation*

$$\Xi_i = \mathbf{W} \begin{bmatrix} r \times r & \mathbf{C}_{11}\mathbf{B}_i^T\mathbf{C}_{22} - \mathbf{C}_{21}^T\mathbf{B}_i\mathbf{C}_{12} \\ \mathbf{C}_{22}\mathbf{B}_i\mathbf{C}_{11} - \mathbf{C}_{12}^T\mathbf{B}_i^T\mathbf{C}_{21} & (p-r) \times (p-r) \end{bmatrix} \mathbf{W}^T,$$

where

$$\mathbf{C}_{\mathbf{A}} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \begin{bmatrix} (\mathbf{I}_r + \mathbf{A}^T\mathbf{A})^{-1} & -(\mathbf{I}_r + \mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \\ \mathbf{A}(\mathbf{I}_r + \mathbf{A}^T\mathbf{A})^{-1} & \mathbf{I}_{p-r} - \mathbf{A}(\mathbf{I}_r + \mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \end{bmatrix},$$

$\mathbf{W} = \mathbf{P}\mathbf{C}_{\mathbf{A}}^{-T}\mathbf{C}_{\mathbf{A}}$, and $\mathbf{B}_i \in \mathbb{R}^{(p-r) \times r}$, $i = 1, 2$. Then the symmetric bilinear form

$$\langle \Xi_1, \Xi_2 \rangle_{\Pi} := \langle \mathbf{C}_{11}\mathbf{B}_1\mathbf{C}_{22} - \mathbf{C}_{12}^T\mathbf{B}_1^T\mathbf{C}_{21}, \mathbf{C}_{11}\mathbf{B}_2\mathbf{C}_{22} - \mathbf{C}_{12}^T\mathbf{B}_2^T\mathbf{C}_{21} \rangle_{\mathbf{F}}$$

is positive definite, where $\langle \cdot, \cdot \rangle_{\mathbf{F}}$ is the Frobenius inner product between matrices.

It is clear that \mathbf{W} defined in Proposition 4.2 plays the same role as \mathbf{Q} , and $\mathbf{Y}_i = \mathbf{C}_{22}\mathbf{B}_i\mathbf{C}_{11} - \mathbf{C}_{12}^T\mathbf{B}_i^T\mathbf{C}_{21}$ under the parameterization $(\mathcal{A}, \mathbf{y}_{\mathbf{P}})$ introduce in (4.1). This implies that the metric defined in Proposition 4.2 coincides with the canonical metric on $\mathcal{G}(p, r)$. Consequently, the connections, gradients, and geodesics can be further derived using the chart (4.1). Below, we provide an example of the geodesics derived through the chart (4.1) when $r = 1$.

EXAMPLE 4.3 (Geodesics on $\mathcal{G}(p, 1)$). *Consider the geodesics on $\mathcal{G}(p, 1)$. Let $\Pi = \mathbf{u}\mathbf{u}^T$ for some $\mathbf{u} \in \mathbb{O}(p, 1)$ and $\Xi \in T_{\Pi}\mathcal{G}(p, 1)$. Then for small enough $\epsilon > 0$, the geodesic $\gamma : (-\epsilon, \epsilon) \rightarrow \mathcal{G}(p, 1)$ with $\gamma(0) = \Pi$ and initial velocity $\gamma'(0) = \Xi$ is given by*

$$\gamma(t) = \left\{ \mathbf{u} \cos(\|\Xi\mathbf{u}\|_2 t) + \frac{\Xi\mathbf{u}}{\|\Xi\mathbf{u}\|_2} \sin(\|\Xi\mathbf{u}\|_2 t) \right\} \times \left\{ \mathbf{u} \cos(\|\Xi\mathbf{u}\|_2 t) + \frac{\Xi\mathbf{u}}{\|\Xi\mathbf{u}\|_2} \sin(\|\Xi\mathbf{u}\|_2 t) \right\}^T.$$

The derivation of the geodesics is deferred to Appendix. Note that our derivation is obtained by solving the geodesic equation in the chart (4.1) directly and differs from those in [3, 8, 26]. The geodesics on $\mathcal{G}(p, r)$ for general $r \geq 2$ can be found

in [3, 26, 58]. For small $t > 0$, it can be shown that the geodesic distance between $\gamma(0) = \mathbf{\Pi}$ and $\gamma(t)$ is given by $d(\mathbf{\Pi}, \gamma(t)) = \sigma t$, where $\sigma := \|\mathbf{\Xi}\mathbf{u}\|_2$ is the canonical angle between \mathbf{u} and $\mathbf{u} \cos \sigma t + (\mathbf{\Xi}\mathbf{u}/\sigma) \sin \sigma t$ (see, for example, Section 3.8 of [3] or Section 2.5.1 of [26]).

We further provide a numerical illustration of the geodesic distance on $\mathcal{G}(p, 1)$ compared to several other distances. Suppose $\mathbf{\Pi} = \mathbf{u}\mathbf{u}^T$ with $p = 200$, $\mathbf{u} = p^{-1/2}\mathbf{1}_p$, $\sigma = \|\mathbf{\Xi}\mathbf{u}\|_2 = 1$, and $\mathbf{\Xi}\mathbf{u} = p^{-1/2}[\mathbf{1}_{p/2}^T, -\mathbf{1}_{p/2}^T]^T$, where $\mathbf{1}_k$ is the vector of all ones in \mathbb{R}^k for $k \geq 1$. It is straightforward to show that $\gamma(t) = \mathbf{y}_I(\varphi(t))$, where

$$\varphi(t) = \frac{\mathbf{1}_{p-1} \cos t + [\mathbf{1}_{p/2-1}^T, -\mathbf{1}_{p/2}^T]^T \sin t}{p^{1/2} + \cos t + \sin t}.$$

Let $\mathbf{u}(t) = \mathbf{u} \cos \sigma t + (\mathbf{\Xi}\mathbf{u}/\sigma) \sin \sigma t$. We plot the geodesic distance $d(\mathbf{u}(t)\mathbf{u}(t)^T, \mathbf{u}\mathbf{u}^T)$ on the Grassmannian $\mathcal{G}(p, 1)$, the Frobenius sine-theta distance $\|\sin \Theta(\mathbf{u}(t), \mathbf{u})\|_F$ between subspaces, the Euclidean distance $\|\varphi(t) - \varphi(0)\|_2$ with regard to $\varphi(t)$, and the Euclidean distance $\|\mathbf{u}(t) - \mathbf{u}\|_2$ with regard to $\mathbf{u}(t)$ as functions of $t \in [0, 1]$. We see that the four distances are quite comparable to each other. This finding also numerically verifies the perturbation bounds established in Section 3.1.

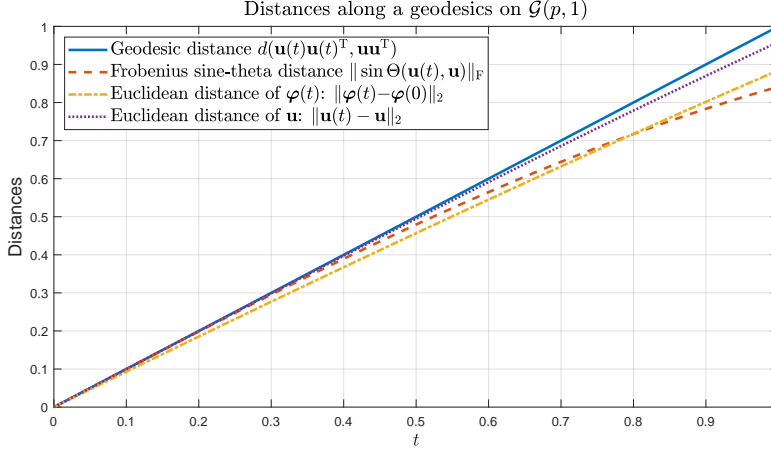


FIG. 1. The geodesic distance $d(\mathbf{u}(t)\mathbf{u}(t)^T, \mathbf{u}\mathbf{u}^T)$ on $\mathcal{G}(p, 1)$, the Frobenius sine-theta distance $\|\sin \Theta(\mathbf{u}(t), \mathbf{u})\|_F$ between subspaces, the Euclidean distance $\|\varphi(t) - \varphi(0)\|_2$ with regard to $\varphi(t)$, and the Euclidean distance $\|\mathbf{u}(t) - \mathbf{u}\|_2$ with regard to $\mathbf{u}(t)$ as functions of $t \in [0, 1]$ for the numerical experiment in Example 4.3.

4.2. Multiplicative perturbation. The intrinsic perturbation theory established in Section 3.1 is particularly suited for the multiplicative perturbation analysis [33, 38, 39], in which case the low-rank structure is always preserved. Multiplicative perturbation theory is also useful in numerical linear algebra applications such as polar decompositions [37, 40], QR factorizations [19], and singular value decompositions [33]. Formally, let $\mathbf{\Sigma}, \mathbf{\Sigma}_0 \in \mathcal{S}(p, r)$ and $\mathbf{\theta}, \mathbf{\theta}_0 \in \mathcal{D}(p, r)$ such that $\mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{\theta})$ and $\mathbf{\Sigma}_0 = \mathbf{\Sigma}(\mathbf{\theta}_0)$, where $\mathcal{S}(p, r)$, $\mathcal{D}(p, r)$, and $\mathbf{\Sigma}(\cdot)$ are defined in Section 2.4. The two matrices $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_0$ are related by a non-singular multiplicative factor $\mathbf{D} \in \mathbb{R}^{p \times p}$ simultaneously from the left and right, namely, $\mathbf{\Sigma} = \mathbf{D}\mathbf{\Sigma}_0\mathbf{D}^T$. The multiplicative perturbation analysis concerns the variation of the eigenspace of $\mathbf{\Sigma}$ relative to that of

Σ_0 as a function of $\mathbf{I}_p - \mathbf{D}$. For simplicity, we consider the case where $r = 1$. Proposition 4.4 below, the proof of which is deferred to Appendix, provides the perturbation results of the leading eigenspace in the context of multiplicative perturbations.

PROPOSITION 4.4. Let $\Sigma(\cdot)$ be the matrix-valued function defined in (2.6) and $\theta, \theta_0 \in \mathcal{D}(p, 1)$, where $\mathcal{D}(p, 1)$ is defined in (2.7). Let $\theta = [\varphi^T, \mu^T]^T$ and $\theta_0 = [\varphi_0^T, \mu_0^T]^T$, where $\varphi, \varphi_0 \in \mathbb{R}^{p-1}$ with $\|\varphi\|_2, \|\varphi_0\|_2 < 1$. Suppose $\Sigma(\theta) = \mathbf{D}\Sigma(\theta_0)\mathbf{D}^T$ for some non-singular $\mathbf{D} \in \mathbb{R}^{p \times p}$ and $\mathbf{D} - \mathbf{I}_p$ has the following block form

$$\mathbf{D} - \mathbf{I}_p = \begin{bmatrix} \epsilon & \alpha^T \\ \beta & \Gamma \end{bmatrix},$$

where $\epsilon \in \mathbb{R}$, $\alpha, \beta \in \mathbb{R}^{p-1}$, and $\Gamma \in \mathbb{R}^{(p-1) \times (p-1)}$. If

$$\|\mathbf{D} - \mathbf{I}_p\|_F \leq \min \left\{ \frac{1}{4}, \frac{1 - \varphi_0^T \varphi_0}{4(1 + \varphi_0^T \varphi_0)} \right\},$$

then

$$\begin{aligned} \|\varphi - \varphi_0\|_2 \leq & \frac{(1 + \varphi_0^T \varphi_0)(\|\mathbf{D}\mathbf{u}_0\|_2 - 1)}{2 - 4\|\mathbf{D} - \mathbf{I}_p\|_F} + \frac{(1 - \varphi_0^T \varphi_0)(|\epsilon| + \|\beta\|_2)}{2 - 4\|\mathbf{D} - \mathbf{I}_p\|_F} \\ & + \frac{2\|\Gamma\varphi_0 - \alpha^T \varphi_0 \varphi_0\|_2}{2 - 4\|\mathbf{D} - \mathbf{I}_p\|_F}. \end{aligned}$$

Proposition 4.4 implies that when the multiplicative matrix \mathbf{D} has a small deviation from the identity matrix \mathbf{I}_p , the perturbation of the corresponding representing Euclidean vectors $\|\varphi - \varphi_0\|_2$ can be controlled by the deviation of \mathbf{D} from \mathbf{I}_p characterized in terms of the blocks ϵ , α , β , and Γ because $|\|\mathbf{D}\mathbf{u}_0\|_2 - 1| \leq \|\mathbf{D} - \mathbf{I}_p\|_2$ and $\|\Gamma\varphi_0 - \alpha^T \varphi_0 \varphi_0\|_2 \leq \|\Gamma\|_2 + \|\alpha\|_2$. Below, Example 4.5 illustrates the above multiplicative perturbation bound compared with the relative perturbation bound in [39] and the classical Davis-Kahan perturbation bound.

EXAMPLE 4.5. Consider the multiplicative perturbation when $r = 1$. Let $p = 200$, $\Sigma_0 = \lambda_0 \mathbf{u}_0 \mathbf{u}_0^T$, $\lambda_0 = 10$, $\mathbf{u}_0 = p^{-1/2} \mathbf{1}_p$, and $\varphi_0 = (p^{1/2} + 1)^{-1} \mathbf{1}_{p-1}$. Generate a multiplicative perturbation matrix $\Delta = [\Delta_{ij}]_{p \times p}$ with Δ_{ij} being independent and identically distributed $N(0, 1)$ random variables and set $\mathbf{D} = \mathbf{I}_p + c\Delta/\|\Delta\|_F$, where $c = \min[1/4, (1 - \varphi_0^T \varphi_0)/\{4(1 + \varphi_0^T \varphi_0)\}]^{-1}$. Let \mathbf{u} be the leading eigenvector of $\Sigma := \mathbf{D}\Sigma_0\mathbf{D}^T$. To find the representing vector $\phi \in \mathbb{R}^{p-1}$ and $\mu \in \mathbb{R}$ such that $\Sigma(\theta) = \mathbf{D}\Sigma_0\mathbf{D}^T$, where $\theta = [\varphi^T, \mu]^T$, note that

$$\mathbf{D}\Sigma_0\mathbf{D}^T = \left(\frac{\mathbf{D}\mathbf{u}_0}{\|\mathbf{D}\mathbf{u}_0\|_2} \right) (\lambda_0 \|\mathbf{D}\mathbf{u}_0\|_2^2) \left(\frac{\mathbf{D}\mathbf{u}_0}{\|\mathbf{D}\mathbf{u}_0\|_2} \right)^T.$$

From the proof of Proposition 4.4, we obtain that $\mu = \lambda_0 \|\mathbf{D}\mathbf{u}_0\|_2^2$ and

$$\varphi = \begin{bmatrix} p-1 & \mathbf{I}_{p-1} \end{bmatrix} \frac{\mathbf{D}\mathbf{u}_0}{\|\mathbf{D}\mathbf{u}_0\|_2} \left(1 + \frac{\mathbf{I}_{p-1}^T \mathbf{D}\mathbf{u}_0}{\|\mathbf{D}\mathbf{u}_0\|_2} \right)^{-1}.$$

In this case, one can construct Σ from $\theta = [\varphi^T, \mu]^T$ using the formula

$$\Sigma = \lambda \begin{bmatrix} \frac{1 - \varphi^T \varphi}{1 + \varphi^T \varphi} \\ 2\varphi \\ 1 + \varphi^T \varphi \end{bmatrix} \begin{bmatrix} \frac{1 - \varphi^T \varphi}{1 + \varphi^T \varphi} & \frac{2\varphi^T}{1 + \varphi^T \varphi} \end{bmatrix}.$$

In [39], the author showed the following relative perturbation bound:

$$\|\sin \Theta(\mathbf{u}, \mathbf{u}_0)\|_F \leq \sqrt{\|(\mathbf{I} - \mathbf{D}^{-1})\mathbf{u}_0\|_F^2 + \|(\mathbf{I} - \mathbf{D})\mathbf{u}_0\|_F^2}.$$

We compare the above relative perturbation bound with the multiplicative perturbation bound obtained in Proposition 4.4 as well as the classical Davis-Kahan perturbation bound numerically using 500 independent Monte Carlo replicates. Figure 2 visualizes the Euclidean distance $\|\varphi - \varphi_0\|_2$, our multiplicative perturbation bound in Proposition 4.4, the sine-theta distance $\|\sin \Theta(\mathbf{u}, \mathbf{u}_0)\|_2$ between subspaces, the relative perturbation bound of [39], and the classical Davis-Kahan perturbation bound $2\|\Sigma - \Sigma_0\|_2/\lambda_0$ (see, for example, Theorem 6.9 in [18]) across 500 independent numerical experiments. It can be seen that the sine-theta distance between subspaces is close to the Euclidean distance between their representing vectors. Also, our multiplicative perturbation bound seems to be sharper than the relative perturbation bound of [39] and the Davis-Kahan perturbation bound because our upper bound takes advantage of both the rank-preserving nature of the multiplicative perturbation $\Sigma_0 \mapsto \mathbf{D}\Sigma_0\mathbf{D}^T$ and the block structure of the multiplicative factor \mathbf{D} .

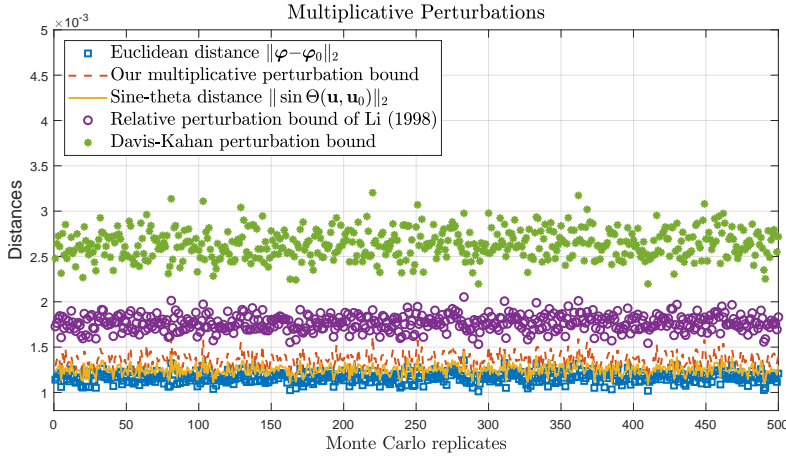


FIG. 2. The Euclidean distance $\|\varphi - \varphi_0\|_2$, our multiplicative perturbation bound in Proposition 4.4, the sine-theta distance $\|\sin \Theta(\mathbf{u}, \mathbf{u}_0)\|_2$, the relative perturbation bound of [39], and the classical Davis-Kahan perturbation bound across 500 independent numerical experiments in Example 4.5.

5. Extension to general rectangular matrices. Using a similar approach, we can extend the Euclidean representation framework for symmetric low-rank matrices to general and possibly rectangular low-rank matrices, which can be applied to a broader range of problems. Rather than using Euclidean vectors to represent the eigenspaces as an intermediate step, we use Euclidean vectors to represent the corresponding right singular subspaces as follows.

Suppose Σ is a $p_1 \times p_2$ matrix with rank r and let $\Sigma = \mathbf{V}_1 \mathbf{\Lambda} \mathbf{V}_2^T$ be its singular value decomposition, where $\mathbf{V}_1 \in \mathbb{O}(p_1, r)$, $\mathbf{V}_2 \in \mathbb{O}(p_2, r)$, and $\mathbf{\Lambda}$ is the diagonal matrix of the singular values of Σ . Assume that $\mathbf{I}_{p_2 \times r}^T \mathbf{V}_2$ is invertible and let $\mathbf{I}_{p_2 \times r}^T \mathbf{V}_2 = \mathbf{W}_1 \text{diag}\{\sigma_1(\mathbf{I}_{p_2 \times r}^T \mathbf{V}_2), \dots, \sigma_r(\mathbf{I}_{p_2 \times r}^T \mathbf{V}_2)\} \mathbf{W}_2^T$ be its singular value decomposition, where $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{O}(r)$. Similar to Section 2.4, we can parameterize the rectangular matrix Σ using the two matrices $\mathbf{U} = \mathbf{V}_2 \mathbf{W}_2 \mathbf{W}_1^T$ and $\mathbf{M} = \mathbf{V}_1 \mathbf{\Lambda} \mathbf{W}_2 \mathbf{W}_1^T$.

Clearly, $\Sigma = \mathbf{M}\mathbf{U}^T$, where $\mathbf{M} \in \mathbb{R}^{p_1 \times r}$ is a constraint-free full-rank matrix and $\mathbf{U} \in \mathbb{O}_+(p_2, r)$. Invoking the Cayley parameterization (2.1), we can further reparameterize \mathbf{U} using a $(p_2 - r) \times r$ matrix \mathbf{A} with $\|\mathbf{A}\|_2 < 1$, such that $\mathbf{U} = \mathbf{U}(\varphi)$, where $\varphi = \text{vec}(\mathbf{A}) \in \mathbb{R}^{(p_2-r)r}$. In particular, we see that $\text{Span}(\mathbf{V}_2) = \text{Span}(\mathbf{U})$. Now denote $\theta = [\varphi^T, \mu^T]^T$, where $\mu = \text{vec}(\mathbf{M})$. Then we can view Σ generically as a matrix-valued function of θ :

$$(5.1) \quad \theta \mapsto \Sigma(\theta) = \mathbf{M}(\mu)\mathbf{U}(\varphi)^T,$$

where $\mathbf{M}(\cdot) : \mathbb{R}^{p_1 r} \rightarrow \mathbb{R}^{p_1 \times r}$ is the inverse of the function $\text{vec}(\cdot) : \mathbb{R}^{p_1 \times r} \rightarrow \mathbb{R}^{p_1 r}$.

We use $\Sigma_0 \in \mathbb{R}^{p_1 \times p_2}$ to denote the referential matrix of interest. Let $\text{rank}(\Sigma_0) = r \leq \min(p_1, p_2)$ and the right singular vector matrix of Σ_0 be \mathbf{V}_2 . We further assume that $\mathbf{I}_{p_2 \times r}^T \mathbf{V}_2$ is non-singular. Similarly, Σ_0 can also be represented by a Euclidean vector θ_0 . Write $\theta_0 = [\varphi_0^T, \mu_0^T]^T \in \mathbb{R}^{(p_2-r)r} \times \mathbb{R}^{p_1 r}$, where $\varphi_0 = \text{vec}(\mathbf{A}_0)$ for a $(p_2 - r) \times r$ matrix \mathbf{A}_0 with $\|\mathbf{A}_0\|_2 < 1$ and $\mu_0 = \text{vec}(\mathbf{M}_0)$ for a $p_1 \times r$ full-rank matrix \mathbf{M}_0 . Let $\mathbf{U}_0 := \mathbf{U}(\varphi_0)$. We then define the following matrix-valued functions, extending the functions in (2.9) to general rectangular $\Sigma(\cdot)$:

$$(5.2) \quad \begin{aligned} D_\varphi \Sigma(\theta) &= \mathbf{K}_{p_2 p_1}(\mathbf{M} \otimes \mathbf{I}_{p_2}) D\mathbf{U}(\varphi), \\ D_\mu \Sigma(\theta) &= \mathbf{U}(\varphi) \otimes \mathbf{I}_{p_1}, \\ D\Sigma(\theta) &= [D_\varphi \Sigma(\theta) \quad D_\mu \Sigma(\theta)]. \end{aligned}$$

Theorem 5.1 below, which extends Theorem 3.1 to general rectangular matrices, asserts that $D\Sigma(\theta_0)$ defined in (5.2) is exactly the Fréchet derivative of the map $\Sigma(\cdot)$ defined by (5.1) evaluated at $\theta = \theta_0$.

THEOREM 5.1. *Under the setup and notations of Sections 2.2 and 5, if $\|\mathbf{A}_0\|_2 < 1$ and $\|\mathbf{A}\|_2 < 1$, then there exists a $p_1 \times p_2$ matrix $\mathbf{R}(\theta, \theta_0)$ depending on θ, θ_0 , such that*

$$\text{vec}\{\Sigma(\theta) - \Sigma_0\} = D\Sigma(\theta_0)(\theta - \theta_0) + \text{vec}\{\mathbf{R}(\theta, \theta_0)\},$$

where $\mathbf{R}(\theta, \theta_0)$ satisfies $\|\mathbf{R}(\theta, \theta_0)\|_F \leq (4 + 8\|\mathbf{M}_0\|_2)\|\theta - \theta_0\|_F^2$ for all θ, θ_0 . In matrix form, the above displayed equation can be written as

$$\begin{aligned} \Sigma(\theta) - \Sigma_0 &= 2\mathbf{M}_0\mathbf{U}_0(\mathbf{I}_{p_2} - \mathbf{X}_{\varphi_0})^{-1}(\mathbf{X}_{\varphi_0} - \mathbf{X}_\varphi)(\mathbf{I}_{p_2} - \mathbf{X}_{\varphi_0})^{-T} \\ &\quad + (\mathbf{M} - \mathbf{M}_0)\mathbf{U}_0^T + \mathbf{R}(\theta, \theta_0). \end{aligned}$$

Proof. First observe that simple algebra leads to the following matrix decomposition

$$\Sigma - \Sigma_0 = \Sigma(\theta) - \Sigma(\theta_0) = (\mathbf{M} - \mathbf{M}_0)\mathbf{U}_0^T + \mathbf{M}_0\{\mathbf{U}(\varphi) - \mathbf{U}_0\}^T + \mathbf{R}_\Sigma(\theta, \theta_0),$$

where the remainder $\mathbf{R}_\Sigma(\theta, \theta_0) = (\mathbf{M} - \mathbf{M}_0)\{\mathbf{U}(\varphi) - \mathbf{U}_0\}^T$ satisfies

$$\|\mathbf{R}_\Sigma(\theta, \theta_0)\|_F \leq \|\mathbf{M} - \mathbf{M}_0\|_F \|\mathbf{U}(\varphi) - \mathbf{U}_0\|_F \leq \frac{1}{2}\|\mu - \mu_0\|_2^2 + \frac{1}{2}\|\mathbf{U}(\varphi) - \mathbf{U}_0\|_F^2.$$

By Theorem 2.1, $\|\mathbf{U}(\varphi) - \mathbf{U}_0\|_F \leq 2\sqrt{2}\|\varphi - \varphi_0\|_2$. Hence, $\|\mathbf{R}_\Sigma(\theta, \theta_0)\|_F \leq 4\|\theta - \theta_0\|_2^2$. Furthermore, using Theorem 2.1 again, we obtain

$$\text{vec}\{\mathbf{U}(\varphi) - \mathbf{U}_0\} = D\mathbf{U}(\varphi_0)(\varphi - \varphi_0) + \text{vec}\{\mathbf{R}_\mathbf{U}(\varphi, \varphi_0)\},$$

831 where $\|\mathbf{R}_U(\varphi, \varphi_0)\|_F \leq 8\|\varphi - \varphi_0\|_2^2$ for all φ, φ_0 . In matrix form, we have

$$832 \quad \mathbf{U}(\varphi) - \mathbf{U}_0 = 2(\mathbf{I}_p - \mathbf{X}_0)^{-1}(\mathbf{X}_\varphi - \mathbf{X}_0)(\mathbf{I}_p - \mathbf{X}_0)^{-1}\mathbf{I}_{p \times r} + \mathbf{R}_U(\varphi, \varphi_0).$$

834 Hence we finally obtain

$$\begin{aligned} 835 \quad \text{vec}\{\Sigma(\theta) - \Sigma_0\} &= D_\mu \Sigma(\theta_0)(\mu - \mu_0) + \mathbf{K}_{p_2 p_1}(\mathbf{M}_0 \otimes \mathbf{I}_{p_2})\text{vec}\{\mathbf{U}(\varphi) - \mathbf{U}_0\} \\ 836 \quad &\quad + \text{vec}\{\mathbf{R}_\Sigma(\theta, \theta_0)\} \\ 837 \quad &= D\Sigma(\theta_0)(\theta - \theta_0) + \text{vec}\{\mathbf{R}(\theta, \theta_0)\}, \end{aligned}$$

839 where $\|\mathbf{R}(\theta, \theta_0)\|_F \leq \|\mathbf{M}_0\|_2 \|\mathbf{R}_U(\varphi, \varphi_0)\|_F + \|\mathbf{R}_\Sigma(\theta, \theta_0)\|_F \leq (4+8\|\mathbf{M}_0\|_2)\|\theta - \theta_0\|_2^2.$

840 Analogously, Theorem 5.2 below also extends Theorem 3.6 to general rectangular
841 $\Sigma(\cdot)$.

842 THEOREM 5.2. *Under the setup and notations of Sections 2.2 and 5,*

$$843 \quad \|\{D\Sigma(\theta_0)^T D\Sigma(\theta_0)\}^{-1}\|_2 \leq \begin{cases} 1 + \frac{(1+8\|\mathbf{M}_0\|_2^2)(1+\|\mathbf{A}_0\|_2^2)^4}{4\sigma_r^2(\mathbf{M}_0)(1-\|\mathbf{A}_0\|_2^2)^2}, & \text{if } r \geq 2, \\ 1 + \frac{(1+8\|\mathbf{M}_0\|_2^2)(1+\|\mathbf{A}_0\|_2^2)^2}{4\sigma_r^2(\mathbf{M}_0)}, & \text{if } r = 1. \end{cases}$$

844 *Proof.* Let $\theta_0 = [\varphi_0^T, \mu_0^T]^T$, where $\varphi_0 = \text{vec}(\mathbf{A}_0)$ for some $\mathbf{A}_0 \in \mathbb{R}^{(p_2-r) \times r}$ and
845 $\mu_0 = \text{vec}(\mathbf{M}_0)$ for some $\mathbf{M}_0 \in \mathbb{R}^{p_1 \times r}$. Let $\mathbf{U}_0 := \mathbf{U}(\varphi_0)$. Then

$$\begin{aligned} 846 \quad D\Sigma(\theta_0)^T D\Sigma(\theta_0) &= \begin{bmatrix} \mathbf{J}_1 & \mathbf{J}_2 \\ \mathbf{J}_2^T & \mathbf{J}_3 \end{bmatrix} := \begin{bmatrix} D_\varphi \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) & D_\varphi \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \\ D_\mu \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) & D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \end{bmatrix} \\ 847 \quad &= \begin{bmatrix} D\mathbf{U}(\varphi)^T(\mathbf{M}_0^T \mathbf{M}_0 \otimes \mathbf{I}_{p_2})D\mathbf{U}(\varphi) & D\mathbf{U}(\varphi)^T(\mathbf{M}_0^T \otimes \mathbf{I}_{p_2})\mathbf{K}_{p_1 p_2}(\mathbf{U}_0 \otimes \mathbf{I}_{p_1}) \\ (\mathbf{U}_0^T \otimes \mathbf{I}_{p_1})\mathbf{K}_{p_2 p_1}(\mathbf{M}_0 \otimes \mathbf{I}_{p_2})D\mathbf{U}(\varphi) & \mathbf{I}_r \otimes \mathbf{I}_{p_1} \end{bmatrix}. \end{aligned}$$

850 The Schur complement of \mathbf{J}_3 of the entire matrix $D\Sigma(\theta_0)^T D\Sigma(\theta_0)$ is given by

$$\begin{aligned} 851 \quad \mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T &= D\mathbf{U}(\varphi_0)^T(\mathbf{M}_0^T \mathbf{M}_0 \otimes \mathbf{I}_{p_2})D\mathbf{U}(\varphi_0) \\ 852 \quad &\quad - D\mathbf{U}(\varphi_0)^T(\mathbf{M}_0^T \otimes \mathbf{I}_{p_2})\mathbf{K}_{p_1 p_2}(\mathbf{U}_0 \mathbf{U}_0^T \otimes \mathbf{I}_{p_1})\mathbf{K}_{p_2 p_1}(\mathbf{M}_0 \otimes \mathbf{I}_{p_2})D\mathbf{U}(\varphi_0). \end{aligned}$$

855 Similar to the proof of Theorem 3.6, we provide a lower bound for the smallest eigen-
856 value of $\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T$. Denote

$$857 \quad \mathbf{X}_0 = \mathbf{X}_{\varphi_0} = \begin{bmatrix} r \times r & -\mathbf{A}_0^T \\ \mathbf{A}_0 & (p_2-r) \times (p_2-r) \end{bmatrix}, \quad \mathbf{C}_0 = (\mathbf{I}_{p_2} - \mathbf{X}_{\varphi_0})^{-1}, \quad \mathbf{W}_0 = \mathbf{C}_0^{-T} \mathbf{C}_0.$$

For any nonzero vector $\varphi \in \mathbb{R}^{(p_2-r)r}$, write

$$\begin{aligned}
& \varphi^T (\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T) \varphi \\
&= \varphi^T D\mathbf{U}(\varphi)^T (\mathbf{M}_0^T \mathbf{M}_0 \otimes \mathbf{I}_{p_2}) D\mathbf{U}(\varphi) \varphi \\
&\quad - \{ \mathbf{K}_{p_2 p_1} (\mathbf{M}_0 \otimes \mathbf{I}_{p_2}) D\mathbf{U}(\varphi_0) \varphi \}^T (\mathbf{U}_0 \mathbf{U}_0^T \otimes \mathbf{I}_{p_1}) \{ \mathbf{K}_{p_2 p_1} (\mathbf{M}_0 \otimes \mathbf{I}_{p_2}) D\mathbf{U}(\varphi_0) \varphi \} \\
&= 4 \{ (\mathbf{I}_{p_2 \times r}^T \mathbf{C}_0^T \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_\varphi) \}^T (\mathbf{M}_0^T \mathbf{M}_0 \otimes \mathbf{I}_{p_2}) \{ (\mathbf{I}_{p_2 \times r}^T \mathbf{C}_0^T \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_\varphi) \} \\
&\quad - 4 \{ \mathbf{K}_{p_2 p_1} (\mathbf{M}_0 \otimes \mathbf{I}_{p_2}) (\mathbf{I}_{p_2 \times r}^T \mathbf{C}_0^T \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_\varphi) \}^T (\mathbf{U}_0 \mathbf{U}_0^T \otimes \mathbf{I}_{p_1}) \\
&\quad \times \{ \mathbf{K}_{p_2 p_1} (\mathbf{M}_0 \otimes \mathbf{I}_{p_2}) (\mathbf{I}_{p_2 \times r}^T \mathbf{C}_0^T \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_\varphi) \} \\
&= 4 \{ \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \}^T (\mathbf{M}_0^T \mathbf{M}_0 \otimes \mathbf{I}_{p_2}) \{ \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \} \\
&\quad - 4 \{ \mathbf{K}_{p_2 p_1} (\mathbf{M}_0 \otimes \mathbf{I}_{p_2}) \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \}^T (\mathbf{U}_0 \mathbf{U}_0^T \otimes \mathbf{I}_{p_1}) \\
&\quad \times \{ \mathbf{K}_{p_2 p_1} (\mathbf{M}_0 \otimes \mathbf{I}_{p_2}) \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \} \\
&= 4 \{ \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \}^T (\mathbf{M}_0^T \mathbf{M}_0 \otimes \mathbf{I}_{p_2}) \{ \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \} \\
&\quad - 4 \{ \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \}^T (\mathbf{M}_0^T \otimes \mathbf{I}_{p_2}) \mathbf{K}_{p_1 p_2} (\mathbf{U}_0 \mathbf{U}_0^T \otimes \mathbf{I}_{p_1}) \mathbf{K}_{p_2 p_1} (\mathbf{M}_0 \otimes \mathbf{I}_{p_2}) \\
&\quad \times \{ \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \} \\
&= 4 \{ \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \}^T (\mathbf{M}_0^T \mathbf{M}_0 \otimes \mathbf{I}_{p_2}) \{ \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \} \\
&\quad - 4 \{ \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \}^T (\mathbf{M}_0^T \mathbf{M}_0 \otimes \mathbf{U}_0 \mathbf{U}_0^T) \\
&\quad \times \{ \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \} \\
&= 4 \{ \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \}^T \{ \mathbf{M}_0^T \mathbf{M}_0 \otimes (\mathbf{I}_{p_2} - \mathbf{U}_0 \mathbf{U}_0^T) \} \{ \text{vec}(\mathbf{C}_0 \mathbf{X}_\varphi \mathbf{C}_0 \mathbf{I}_{p_2 \times r}) \} \\
&= 4 \text{vec}(\mathbf{X}_\varphi) (\mathbf{C}_0^T \otimes \mathbf{C}_0^T) \{ \mathbf{U}_0 \mathbf{M}_0^T \mathbf{M}_0 \mathbf{U}_0^T \otimes (\mathbf{I}_{p_2} - \mathbf{U}_0 \mathbf{U}_0^T) \} (\mathbf{C}_0 \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_\varphi).
\end{aligned}$$

Let $\mathbf{M}_0^T \mathbf{M}_0 = \mathbf{V} \mathbf{S} \mathbf{V}^T$ be the spectral decomposition of $\mathbf{M}_0^T \mathbf{M}_0$, where $\mathbf{V} \in \mathbb{O}(r)$, and \mathbf{S} is the diagonal matrix of the eigenvalues of $\mathbf{M}_0^T \mathbf{M}_0$. Let $\tilde{\mathbf{V}} = \text{diag}(\mathbf{V}, \mathbf{I}_{p_2-r})$. Then we compute

$$\begin{aligned}
& \mathbf{U}_0 \mathbf{M}_0^T \mathbf{M}_0 \mathbf{U}_0^T \otimes (\mathbf{I}_{p_2} - \mathbf{U}_0 \mathbf{U}_0^T) \\
&= \mathbf{W}_0 \tilde{\mathbf{V}} \begin{bmatrix} \mathbf{S} & \\ & \mathbf{I}_{p_2-r} \end{bmatrix} \tilde{\mathbf{V}}^T \mathbf{W}_0^T \otimes \mathbf{W}_0 \tilde{\mathbf{V}} \begin{bmatrix} \mathbf{I}_{p_2-r} & \\ & \mathbf{I}_{p_2-r} \end{bmatrix} \tilde{\mathbf{V}}^T \mathbf{W}_0^T \\
&= (\mathbf{W}_0 \otimes \mathbf{W}_0) (\tilde{\mathbf{V}} \otimes \tilde{\mathbf{V}}) \begin{bmatrix} \mathbf{S} \otimes \begin{bmatrix} \mathbf{I}_{p_2-r} & \\ & \mathbf{I}_{p_2-r} \end{bmatrix} & \\ & \mathbf{I}_{p_2-r} \otimes \mathbf{I}_{p_2-r} \end{bmatrix} \\
&\quad \times (\tilde{\mathbf{V}} \otimes \tilde{\mathbf{V}})^T (\mathbf{W}_0 \otimes \mathbf{W}_0) \\
&\succeq \sigma_r^2(\mathbf{M}_0) (\mathbf{W}_0 \otimes \mathbf{W}_0) \begin{bmatrix} \mathbf{V} \otimes \tilde{\mathbf{V}} & \\ & \mathbf{I}_{p_2-r} \otimes \tilde{\mathbf{V}} \end{bmatrix} \\
&\quad \times \begin{bmatrix} \mathbf{I}_r \otimes \begin{bmatrix} \mathbf{I}_{p_2-r} & \\ & \mathbf{I}_{p_2-r} \end{bmatrix} & \\ & \mathbf{I}_{p_2-r} \otimes \mathbf{I}_{p_2-r} \end{bmatrix} \begin{bmatrix} \mathbf{V}^T \otimes \tilde{\mathbf{V}}^T & \\ & \mathbf{I}_{p_2-r} \otimes \tilde{\mathbf{V}}^T \end{bmatrix} \\
&\quad \times (\mathbf{W}_0^T \otimes \mathbf{W}_0^T) \\
&= \sigma_r^2(\mathbf{M}_0) (\mathbf{W}_0 \otimes \mathbf{W}_0) \begin{bmatrix} \mathbf{I}_r & \\ & \mathbf{I}_{p_2-r} \end{bmatrix} \otimes \begin{bmatrix} \mathbf{I}_{p_2-r} & \\ & \mathbf{I}_{p_2-r} \end{bmatrix} (\mathbf{W}_0^T \otimes \mathbf{W}_0^T) \\
&= \sigma_r^2(\mathbf{M}_0) \{ \mathbf{U}_0 \mathbf{U}_0^T \otimes (\mathbf{I}_{p_2} - \mathbf{U}_0 \mathbf{U}_0^T) \}.
\end{aligned}$$

Therefore, the quadratic form $\varphi^T(\mathbf{J}_1 - \mathbf{J}_2\mathbf{J}_3^{-1}\mathbf{J}_2^T)\varphi$ can be lower bounded:

$$\begin{aligned} & \varphi^T(\mathbf{J}_1 - \mathbf{J}_2\mathbf{J}_3^{-1}\mathbf{J}_2^T)\varphi \\ & \geq 4\sigma_r^2(\mathbf{M}_0) \|(\mathbf{U}_0^T \mathbf{C}_0 \otimes \mathbf{U}_{0\perp}^T \mathbf{C}_0) \text{vec}(\mathbf{X}_\varphi)\|_2^2 \\ & = 4\sigma_r^2(\mathbf{M}_0) \left\| \left\{ (\mathbf{I}_{p_2 \times r}^T \mathbf{C}_0^T \mathbf{C}_0^{-1} \mathbf{C}_0) \otimes [_{(p_2-r) \times r} \quad \mathbf{I}_{p_2-r}] \mathbf{C}_0^T \mathbf{C}_0^{-1} \mathbf{C}_0 \right\} \text{vec}(\mathbf{X}_\varphi) \right\|_2^2 \\ & = 4\sigma_r^2(\mathbf{M}_0) \left\| \left\{ (\mathbf{I}_{p_2 \times r}^T \mathbf{C}_0^T) \otimes [_{(p_2-r) \times r} \quad \mathbf{I}_{p_2-r}] \mathbf{C}_0^T \right\} \text{vec}(\mathbf{X}_\varphi) \right\|_2^2, \end{aligned}$$

where $\mathbf{U}_{0\perp}$ is the orthogonal complement of \mathbf{U}_0 , i.e., $\mathbf{U}_{0\perp} = \mathbf{C}_0^{-T} \mathbf{C}_0 \boldsymbol{\Theta}_2^T$, and $\boldsymbol{\Theta}_2 = [_{(p_2-r) \times r}, \mathbf{I}_{p_2-r}]$. Write \mathbf{C}_0 in the block matrix form

$$\mathbf{C}_0 = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \begin{bmatrix} (\mathbf{I}_r + \mathbf{A}_0^T \mathbf{A}_0)^{-1} & -(\mathbf{I}_r + \mathbf{A}_0^T \mathbf{A}_0)^{-1} \mathbf{A}_0^T \\ \mathbf{A}_0(\mathbf{I}_r + \mathbf{A}_0^T \mathbf{A}_0)^{-1} & \mathbf{I}_{p_2-r} - \mathbf{A}_0(\mathbf{I}_r + \mathbf{A}_0^T \mathbf{A}_0)^{-1} \mathbf{A}_0^T \end{bmatrix}.$$

It follows that

$$\begin{aligned} \varphi^T(\mathbf{J}_1 - \mathbf{J}_2\mathbf{J}_3^{-1}\mathbf{J}_2^T)\varphi & \geq 4\sigma_r^2(\mathbf{M}_0) \left\| \left\{ \begin{bmatrix} \mathbf{C}_{11}^T & \mathbf{C}_{21}^T \end{bmatrix} \otimes \begin{bmatrix} \mathbf{C}_{12}^T & \mathbf{C}_{22}^T \end{bmatrix} \right\} \text{vec}(\mathbf{X}_\varphi) \right\|_2^2 \\ & = 4\sigma_r^2(\mathbf{M}_0) \left\| \begin{bmatrix} \mathbf{C}_{12}^T & \mathbf{C}_{22}^T \end{bmatrix} \begin{bmatrix} r \times r & -\mathbf{A}^T \\ \mathbf{A} & (p_2-r) \times (p_2-r) \end{bmatrix} \begin{bmatrix} \mathbf{C}_{11} \\ \mathbf{C}_{21} \end{bmatrix} \right\|_F^2 \\ & = 4\sigma_r^2(\mathbf{M}_0) \left\| \mathbf{C}_{22}^T \mathbf{A} \mathbf{C}_{11} - \mathbf{C}_{12}^T \mathbf{A}^T \mathbf{C}_{21} \right\| \\ & = 4\sigma_r^2(\mathbf{M}_0) \left\| \{(\mathbf{C}_{11} \otimes \mathbf{C}_{22}) - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T) \mathbf{K}_{(p_2-r)r}\} \text{vec}(\mathbf{A}) \right\|_F^2. \end{aligned}$$

We finally invoke Lemma 3.8 to conclude that

$$\varphi^T(\mathbf{J}_1 - \mathbf{J}_2\mathbf{J}_3^{-1}\mathbf{J}_2^T)\varphi \geq \begin{cases} \frac{4\sigma_r^2(\mathbf{M}_0)(1 - \|\mathbf{A}_0\|_2^2)^2 \|\varphi\|_2^2}{(1 + \|\mathbf{A}_0\|_2^2)^4}, & \text{if } r \geq 2, \\ \frac{4\sigma_r^2(\mathbf{M}_0) \|\varphi\|_2^2}{(1 + \|\mathbf{A}_0\|_2^2)^2}, & \text{if } r = 1, \end{cases}$$

which further implies that

$$\|(\mathbf{J}_1 - \mathbf{J}_2\mathbf{J}_3^{-1}\mathbf{J}_2^T)^{-1}\|_2 \leq \begin{cases} \frac{(1 + \|\mathbf{A}_0\|_2^2)^4}{4\sigma_r^2(\mathbf{M}_0)(1 - \|\mathbf{A}_0\|_2^2)^2}, & \text{if } r \geq 2, \\ \frac{(1 + \|\mathbf{A}_0\|_2^2)^2}{4\sigma_r^2(\mathbf{M}_0)}, & \text{if } r = 1. \end{cases}$$

This shows that $\mathbf{J}_1 - \mathbf{J}_2\mathbf{J}_3^{-1}\mathbf{J}_2^T$ is invertible. Since \mathbf{J}_3 is also invertible, the property of the Schur complement immediately implies that $D\Sigma(\boldsymbol{\theta}_0)^T D\Sigma(\boldsymbol{\theta}_0)$ is invertible. Furthermore, by the block matrix inversion formula,

$$\begin{aligned} & \{D\Sigma(\boldsymbol{\theta}_0)^T D\Sigma(\boldsymbol{\theta}_0)\}^{-1} \\ & = \begin{bmatrix} (\mathbf{J}_1 - \mathbf{J}_2\mathbf{J}_3^{-1}\mathbf{J}_2^T)^{-1} & -(\mathbf{J}_1 - \mathbf{J}_2\mathbf{J}_3^{-1}\mathbf{J}_2^T)^{-1} \mathbf{J}_2 \mathbf{J}_3^{-1} \\ -\mathbf{J}_3^{-1} \mathbf{J}_2^T (\mathbf{J}_1 - \mathbf{J}_2\mathbf{J}_3^{-1}\mathbf{J}_2^T)^{-1} & \mathbf{J}_3^{-1} + \mathbf{J}_3^{-1} \mathbf{J}_2^T (\mathbf{J}_1 - \mathbf{J}_2\mathbf{J}_3^{-1}\mathbf{J}_2^T)^{-1} \mathbf{J}_2 \mathbf{J}_3^{-1} \end{bmatrix}. \end{aligned}$$

By construction, $\|\mathbf{J}_3^{-1}\|_2 = 1$ and

$$\begin{aligned} \|\mathbf{J}_2\|_2 & \leq \|D\mathbf{U}(\varphi_0)^T (\mathbf{M}_0^T \otimes \mathbf{I}_p) \mathbf{K}_{p_2 p_1} (\mathbf{U}_0 \otimes \mathbf{I}_{p_1})\| \leq \|D\mathbf{U}(\varphi_0)\|_2 \|\mathbf{M}_0\|_2 \\ & \leq 2\sqrt{2} \|\mathbf{M}_0\|_2 \|\mathbf{I}_{p \times r} (\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\|_2 \|(\mathbf{I}_p - \mathbf{X}_{\varphi_0})^{-1}\|_2 \leq 2\sqrt{2} \|\mathbf{M}_0\|_2. \end{aligned}$$

Thus, by Lemma 3.4 of [10], we see that

$$\begin{aligned}
& \|\{D\Sigma(\theta_0)^T D\Sigma(\theta_0)\}^{-1}\|_2 \leq \|(\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1}\|_2 \\
& \quad + \|\mathbf{J}_3^{-1} + \mathbf{J}_3^{-1} \mathbf{J}_2^T (\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1} \mathbf{J}_2 \mathbf{J}_3^{-1}\|_2 \\
& \leq \|(\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1}\|_2 \\
& \quad + \|\mathbf{J}_3^{-1}\|_2 + \|\mathbf{J}_3^{-1}\|_2^2 \|\mathbf{J}_2\|_2^2 \|(\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1}\|_2 \\
& \leq 1 + (1 + 8\|\mathbf{M}_0\|_2^2) \|(\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1}\|_2.
\end{aligned}$$

The proof is completed by combining the upper bound for $\|(\mathbf{J}_1 - \mathbf{J}_2 \mathbf{J}_3^{-1} \mathbf{J}_2^T)^{-1}\|_2$. \square

6. Conclusion. In this paper, we present a novel Euclidean representation for low-rank matrices and, correspondingly, develop a collection of technical devices for studying the intrinsic perturbation of low-rank matrices, where the referential matrix and the perturbed matrix have the same rank. Our analyses establish that the Frobenius distance between low-rank matrices is locally equivalent to the Euclidean distance between their representing vectors. We also establish that the representation function for low-rank matrices is regular in the sense that its Fréchet derivative (Jacobian) has full column rank, thereby showing that the space of low-rank matrices is a manifold and the representation function leads to a chart for this manifold. Two applications are showcased, namely, the construction of the Riemannian metric on the Grassmannian using the Cayley parameterization and the multiplicative perturbation analysis of low-rank matrices.

These technical tools can potentially apply to a broad range of concrete low-rank matrix problems. For example, in Bayesian statistics, the referential matrix Σ_0 may correspond to the ground truth of the parameter of interest, and Σ is, under the posterior distribution, a random matrix taking values in a low-rank matrix manifold such that $\text{rank}(\Sigma) = \text{rank}(\Sigma_0)$. Many specific statistical models are also built upon low-rank matrix structures, such as principal component analysis, multidimensional scaling, high-dimensional mixture models, stochastic block models and their offspring, and canonical correlation analysis. We defer the application of the present framework to specific high-dimensional statistics problems to future research topics.

There are certain limitations of the theory currently in the paper. As mentioned in Section 1, one example is the entrywise eigenvector and singular vector perturbation analysis with delocalized eigenvectors, which has been attracting growing interest recently in random graph inference and random matrix theory (see, for example, [2, 5, 15, 16, 17, 18, 59]). In short, a delocalized eigenvector $\mathbf{U} \in \mathbb{O}(p, r)$ has the maximum row-wise Euclidean norm that can be upper bounded by a constant factor of $\sqrt{r/p}$. Under the Cayley parameterization $\varphi = \text{vec}(\mathbf{A}) \mapsto \mathbf{U} = \mathbf{U}(\varphi)$, this requires that $\|\mathbf{A}\|_2$ is close to 1, and the resulting Jacobian matrix $D\mathbf{U}(\varphi)$ becomes ill-conditioned. Hence, when the entrywise eigenvector or singular vector analysis is of interest, it is more preferred that the classical extrinsic perturbation tools such as those developed in [2, 5, 18, 59] are applied.

Appendix: Technical Proofs.

Proof of Lemma 3.3. For convenience we denote $\mathbf{U} = \mathbf{U}(\varphi)$ and $\mathbf{U}_0 = \mathbf{U}(\varphi_0)$. First note that

$$\begin{aligned}
\|\mathbf{U}\mathbf{U}^T - \mathbf{U}_0\mathbf{U}_0^T\|_F^2 &= \left\| \begin{bmatrix} \mathbf{Q}_1^2 - \mathbf{Q}_{01}^2 & \mathbf{Q}_1\mathbf{Q}_2^T - \mathbf{Q}_{01}\mathbf{Q}_{02}^T \\ \mathbf{Q}_2\mathbf{Q}_1 - \mathbf{Q}_{02}\mathbf{Q}_{01} & \mathbf{Q}_2\mathbf{Q}_2^T - \mathbf{Q}_{02}\mathbf{Q}_{02}^T \end{bmatrix} \right\|_F^2 \\
&= \|\mathbf{Q}_1^2 - \mathbf{Q}_{01}^2\|_F^2 + 2\|\mathbf{Q}_2\mathbf{Q}_1 - \mathbf{Q}_{02}\mathbf{Q}_{01}\|_F^2 + \|\mathbf{Q}_2\mathbf{Q}_2^T - \mathbf{Q}_{02}\mathbf{Q}_{02}^T\|_F^2.
\end{aligned}$$

For any positive definite matrix $\mathbf{Q} \in \mathbb{R}^{r \times r}$, denote $\Theta(\mathbf{Q}) = \mathbf{Q}^{1/2}$ the matrix square root function evaluated at \mathbf{Q} . Then the matrix differential technique yields

$$\frac{\partial \text{vec}\{\Theta(\mathbf{Q})\}}{\partial \text{vec}(\mathbf{Q})^T} = (\mathbf{Q}^{1/2} \otimes \mathbf{I}_r + \mathbf{I}_r \otimes \mathbf{Q}^{1/2})^{-1} \preceq \frac{1}{2\sigma_{\min}(\mathbf{Q}^{1/2})} \mathbf{I}_{r^2} = \frac{1}{2\sigma_{\min}^{1/2}(\mathbf{Q})} \mathbf{I}_{r^2}.$$

By the mean-value inequality,

$$\begin{aligned} \|\mathbf{Q}_1 - \mathbf{Q}_{01}\|_F &= \|\Theta(\mathbf{Q}_1^2) - \Theta(\mathbf{Q}_{01}^2)\|_F \\ &\leq \max_{t \in [0,1]} \left\| \frac{\partial \text{vec}\{\Theta(\mathbf{Q})\}}{\partial \text{vec}(\mathbf{Q})^T} \right\|_{\mathbf{Q}=(t\mathbf{Q}_{01}^2+(1-t)\mathbf{Q}_1^2)} \|\mathbf{Q}_1^2 - \mathbf{Q}_{01}^2\|_F \\ &= \frac{1}{2} \max_{t \in [0,1]} \lambda_{\min}^{-1/2}\{t\mathbf{Q}_{01}^2 + (1-t)\mathbf{Q}_1^2\} \|\mathbf{Q}_1^2 - \mathbf{Q}_{01}^2\|_F \\ &\leq \frac{\|\mathbf{Q}_1^2 - \mathbf{Q}_{01}^2\|_F}{2 \min\{\lambda_r(\mathbf{Q}_{01}), \lambda_r(\mathbf{Q}_1)\}} \\ &\leq \frac{\|\mathbf{U}\mathbf{U}^T - \mathbf{U}_0\mathbf{U}_0^T\|_F}{2 \min\{\lambda_r(\mathbf{Q}_{01}), \lambda_r(\mathbf{Q}_1)\}}. \end{aligned}$$

For \mathbf{Q}_2 , we have

$$\begin{aligned} \mathbf{Q}_2 - \mathbf{Q}_{02} &= \mathbf{Q}_2\mathbf{Q}_{01}\mathbf{Q}_{01}^{-1} - \mathbf{Q}_2\mathbf{Q}_1\mathbf{Q}_{01}^{-1} + \mathbf{Q}_2\mathbf{Q}_1\mathbf{Q}_{01}^{-1} - \mathbf{Q}_{02}\mathbf{Q}_{01}\mathbf{Q}_{01}^{-1} \\ &= \mathbf{Q}_2(\mathbf{Q}_{01} - \mathbf{Q}_1)\mathbf{Q}_{01}^{-1} + (\mathbf{Q}_2\mathbf{Q}_1 - \mathbf{Q}_{02}\mathbf{Q}_{01})\mathbf{Q}_{01}^{-1}. \end{aligned}$$

Using the previous result, we further write

$$\begin{aligned} \|\mathbf{Q}_2 - \mathbf{Q}_{02}\|_F &\leq \|\mathbf{Q}_2\|_2 \|\mathbf{Q}_{01} - \mathbf{Q}_1\|_F \|\mathbf{Q}_{01}^{-1}\|_2 + \|\mathbf{Q}_2\mathbf{Q}_1 - \mathbf{Q}_{02}\mathbf{Q}_{01}\|_F \|\mathbf{Q}_{01}^{-1}\|_2 \\ &\leq \frac{\|\mathbf{U}\mathbf{U}^T - \mathbf{U}_0\mathbf{U}_0^T\|_F \|\mathbf{Q}_{01}^{-1}\|_2}{2 \min\{\lambda_r(\mathbf{Q}_{01}), \lambda_r(\mathbf{Q}_1)\}} + \|\mathbf{U}\mathbf{U}^T - \mathbf{U}_0\mathbf{U}_0^T\|_F \|\mathbf{Q}_{01}^{-1}\|_2. \end{aligned}$$

Hence,

$$\begin{aligned} \|\mathbf{U} - \mathbf{U}_0\|_F &= (\|\mathbf{Q}_1 - \mathbf{Q}_{01}\|_F^2 + \|\mathbf{Q}_2 - \mathbf{Q}_{02}\|_F^2)^{1/2} \leq \|\mathbf{Q}_1 - \mathbf{Q}_{01}\|_F + \|\mathbf{Q}_2 - \mathbf{Q}_{02}\|_F \\ &\leq \frac{\|\mathbf{U}\mathbf{U}^T - \mathbf{U}_0\mathbf{U}_0^T\|_F \|\mathbf{Q}_{01}^{-1}\|_2}{\min\{\lambda_r(\mathbf{Q}_{01}), \lambda_r(\mathbf{Q}_1)\}} + \|\mathbf{U}\mathbf{U}^T - \mathbf{U}_0\mathbf{U}_0^T\|_F \|\mathbf{Q}_{01}^{-1}\|_2 \\ &\leq \frac{2\|\mathbf{U}\mathbf{U}^T - \mathbf{U}_0\mathbf{U}_0^T\|_F \|\mathbf{Q}_{01}^{-1}\|_2}{\min\{\lambda_r(\mathbf{Q}_{01}), \lambda_r(\mathbf{Q}_1)\}}. \end{aligned}$$

This completes the proof of the first inequality. For the second inequality, note that

$$\begin{aligned} &\|(\mathbf{I}_r + \mathbf{Q}_1)^{-1} - (\mathbf{I}_r + \mathbf{Q}_{01})^{-1}\|_F \\ &= \|(\mathbf{I}_r + \mathbf{Q}_1)^{-1}\{(\mathbf{I}_r + \mathbf{Q}_{01}) - (\mathbf{I}_r + \mathbf{Q}_1)\}(\mathbf{I}_r + \mathbf{Q}_{01})^{-1}\|_F \\ &\leq \|(\mathbf{I}_r + \mathbf{Q}_1)^{-1}\|_2 \|\mathbf{Q}_1 - \mathbf{Q}_{01}\|_F \|(\mathbf{I}_r + \mathbf{Q}_{01})^{-1}\|_2 \\ &\leq \|\mathbf{Q}_1 - \mathbf{Q}_{01}\|_F. \end{aligned}$$

Therefore, by Theorem 2.2, $\|\varphi - \varphi_0\|_2 = \|\mathbf{A} - \mathbf{A}_0\|_F \leq 2\|\mathbf{U} - \mathbf{U}_0\|_F$. The proof is completed by combining the obtained upper bound for $\|\mathbf{U} - \mathbf{U}_0\|_F$. \square

999 *Proof of Lemma 3.7.* let $\mathbf{U}_{0\perp}$ be the orthogonal complement of \mathbf{U}_0 such that
 1000 $\mathbf{W}_0 := [\mathbf{U}_0, \mathbf{U}_{0\perp}] \in \mathbb{O}(p)$, and denote

$$1001 \quad \widetilde{\mathbf{M}}_0 = \mathbf{I}_{p \times r} \mathbf{M}_0 \mathbf{I}_{p \times r}^T = \begin{bmatrix} \mathbf{M}_0 & \\ & \mathbf{I}_{(p-r) \times (p-r)} \end{bmatrix}.$$

1002 Clearly, one can take $\mathbf{W}_0 = \mathbf{C}_0^{-T} \mathbf{C}_0$. Suppose $\mathbf{M}_0 = \mathbf{V}_0 \mathbf{\Lambda}_0 \mathbf{V}_0^T$ is the spectral decom-
 1003 position of \mathbf{M}_0 . Then $\widetilde{\mathbf{M}}_0 = \mathbf{I}_{p \times r} \mathbf{V}_0 \mathbf{\Lambda}_0 \mathbf{V}_0^T \mathbf{I}_{p \times r}^T = \widetilde{\mathbf{V}}_0 \widetilde{\mathbf{\Lambda}}_0 \widetilde{\mathbf{V}}_0^T$, where

$$1004 \quad \widetilde{\mathbf{V}}_0 := \begin{bmatrix} \mathbf{V}_0 & \\ & \mathbf{I}_{p-r} \end{bmatrix}, \quad \text{and} \quad \widetilde{\mathbf{\Lambda}}_0 = \begin{bmatrix} \mathbf{\Lambda}_0 & \\ & \mathbf{I}_{(p-r) \times (p-r)} \end{bmatrix}.$$

1005 For convenience let $\lambda_{0k} := \lambda_k(\mathbf{M}_0)$, $k \in [r]$. For the first assertion, write

$$\begin{aligned} 1006 \quad & \Sigma_0^2 \otimes \mathbf{I}_p - 2\Sigma_0 \otimes \Sigma_0 + \mathbf{I}_p \otimes \Sigma_0^2 \\ 1007 \quad & = (\mathbf{W}_0 \otimes \mathbf{W}_0)(\widetilde{\mathbf{M}}_0^2 \otimes \mathbf{I}_p - 2\widetilde{\mathbf{M}}_0 \otimes \widetilde{\mathbf{M}}_0 + \mathbf{I}_p \otimes \widetilde{\mathbf{M}}_0^2)(\mathbf{W}_0^T \otimes \mathbf{W}_0^T) \\ 1008 \quad & = (\mathbf{W}_0 \otimes \mathbf{W}_0) \begin{bmatrix} \mathbf{M}_0^2 \otimes \mathbf{I}_p - 2\mathbf{M}_0 \otimes \widetilde{\mathbf{M}}_0 + \mathbf{I}_r \otimes \widetilde{\mathbf{M}}_0^2 & \\ & \mathbf{I}_{(p-r)} \otimes \widetilde{\mathbf{M}}_0^2 \end{bmatrix} (\mathbf{W}_0^T \otimes \mathbf{W}_0^T) \\ 1009 \quad & = (\mathbf{W}_0 \otimes \mathbf{W}_0)(\widetilde{\mathbf{V}}_0 \otimes \widetilde{\mathbf{V}}_0) \begin{bmatrix} \mathbf{\Lambda}_0^2 \otimes \mathbf{I}_p - 2\mathbf{\Lambda}_0 \otimes \widetilde{\mathbf{\Lambda}}_0 + \mathbf{I}_r \otimes \widetilde{\mathbf{\Lambda}}_0^2 & \\ & \mathbf{I}_{(p-r)} \otimes \widetilde{\mathbf{\Lambda}}_0^2 \end{bmatrix} \\ 1010 \quad & \times (\widetilde{\mathbf{V}}_0^T \otimes \widetilde{\mathbf{V}}_0^T)(\mathbf{W}_0^T \otimes \mathbf{W}_0^T) \\ 1011 \quad & = (\mathbf{W}_0 \otimes \mathbf{W}_0)(\widetilde{\mathbf{V}}_0 \otimes \widetilde{\mathbf{V}}_0) \begin{bmatrix} \lambda_{01}\mathbf{I}_p - \widetilde{\mathbf{\Lambda}}_0 & & & \\ & \ddots & & \\ & & \lambda_{0r}\mathbf{I}_p - \widetilde{\mathbf{\Lambda}}_0 & \\ & & & \widetilde{\mathbf{\Lambda}}_0 & \\ & & & & \ddots & \\ & & & & & \widetilde{\mathbf{\Lambda}}_0 \end{bmatrix}^2 \\ 1012 \quad & \times (\widetilde{\mathbf{V}}_0^T \otimes \widetilde{\mathbf{V}}_0^T)(\mathbf{W}_0^T \otimes \mathbf{W}_0^T). \end{aligned}$$

1014 Since for each $k \in [r]$,

$$1015 \quad (\lambda_{0k}\mathbf{I}_p - \widetilde{\mathbf{\Lambda}}_0)^2 \succeq \lambda_{0r}^2 \begin{bmatrix} r \times r & \\ & \mathbf{I}_{p-r} \end{bmatrix}$$

1017 and

$$1018 \quad \widetilde{\mathbf{\Lambda}}_0^2 \succeq \lambda_{0r}^2 \begin{bmatrix} \mathbf{I}_r & \\ & \mathbf{I}_{(p-r) \times (p-r)} \end{bmatrix},$$

1019 we can further write

$$\begin{aligned} 1020 \quad & \Sigma_0^2 \otimes \mathbf{I}_p - 2\Sigma_0 \otimes \Sigma_0 + \mathbf{I}_p \otimes \Sigma_0^2 \\ 1021 \quad & \succeq \lambda_{0r}^2 (\mathbf{W}_0 \otimes \mathbf{W}_0)(\widetilde{\mathbf{V}}_0 \otimes \widetilde{\mathbf{V}}_0) \begin{bmatrix} \mathbf{I}_r \otimes \begin{bmatrix} r \times r & \\ & \mathbf{I}_{p-r} \end{bmatrix} & \\ & \mathbf{I}_{p-r} \otimes \begin{bmatrix} \mathbf{I}_r & \\ & \mathbf{I}_{(p-r) \times (p-r)} \end{bmatrix} \end{bmatrix} \\ 1022 \quad & \times (\widetilde{\mathbf{V}}_0^T \otimes \widetilde{\mathbf{V}}_0^T)(\mathbf{W}_0^T \otimes \mathbf{W}_0^T). \end{aligned}$$

1024 Now we focus on the matrix on the right-hand side of the previous display. Write

$$\begin{aligned}
1025 & (\tilde{\mathbf{V}}_0 \otimes \tilde{\mathbf{V}}_0) \begin{bmatrix} \mathbf{I}_r \otimes \begin{bmatrix} r \times r & \\ & \mathbf{I}_{p-r} \end{bmatrix} & \\ & \mathbf{I}_{p-r} \otimes \begin{bmatrix} \mathbf{I}_r & \\ & (p-r) \times (p-r) \end{bmatrix} \end{bmatrix} (\tilde{\mathbf{V}}_0^T \otimes \tilde{\mathbf{V}}_0^T) \\
1026 & = \begin{bmatrix} \mathbf{V}_0 \otimes \tilde{\mathbf{V}}_0 & \\ & \mathbf{I}_{p-r} \otimes \tilde{\mathbf{V}}_0 \end{bmatrix} \begin{bmatrix} \mathbf{I}_r \otimes \begin{bmatrix} \mathbf{I}_r & \\ & \mathbf{I}_{p-r} \end{bmatrix} & \\ & \mathbf{I}_{p-r} \otimes \begin{bmatrix} \mathbf{I}_r & \\ & \end{bmatrix} \end{bmatrix} \\
1027 & \times \begin{bmatrix} \mathbf{V}_0^T \otimes \tilde{\mathbf{V}}_0^T & \\ & \mathbf{I}_{p-r} \otimes \tilde{\mathbf{V}}_0^T \end{bmatrix} \\
1028 & = \begin{bmatrix} \mathbf{V}_0 \mathbf{V}_0^T \otimes \begin{bmatrix} \mathbf{V}_0 \mathbf{V}_0^T & \\ & \mathbf{I}_{p-r} \mathbf{I}_{p-r} \mathbf{I}_{p-r} \end{bmatrix} & \\ & \mathbf{I}_{p-r} \otimes \begin{bmatrix} \mathbf{V}_0 \mathbf{I}_r \mathbf{V}_0^T & \\ & \mathbf{I}_{p-r} \mathbf{I}_{p-r} \end{bmatrix} \end{bmatrix} \\
1029 & = \begin{bmatrix} \mathbf{I}_r \otimes \begin{bmatrix} r \times r & \\ & \mathbf{I}_{p-r} \end{bmatrix} & \\ & \mathbf{I}_{p-r} \otimes \begin{bmatrix} \mathbf{I}_r & \\ & (p-r) \times (p-r) \end{bmatrix} \end{bmatrix}. \\
1030 &
\end{aligned}$$

1031 Therefore,

$$\begin{aligned}
1032 & \Sigma_0^2 \otimes \mathbf{I}_p - 2\Sigma_0 \otimes \Sigma_0 + \mathbf{I}_p \otimes \Sigma_0^2 \\
1033 & \succeq \lambda_{0r}^2 (\mathbf{W}_0 \otimes \mathbf{W}_0) (\tilde{\mathbf{V}}_0 \otimes \tilde{\mathbf{V}}_0) \begin{bmatrix} \mathbf{I}_r \otimes \begin{bmatrix} r \times r & \\ & \mathbf{I}_{p-r} \end{bmatrix} & \\ & \mathbf{I}_{p-r} \otimes \begin{bmatrix} \mathbf{I}_r & \\ & (p-r) \times (p-r) \end{bmatrix} \end{bmatrix} \\
1034 & \times (\tilde{\mathbf{V}}_0^T \otimes \tilde{\mathbf{V}}_0^T) (\mathbf{W}_0^T \otimes \mathbf{W}_0^T) \\
1035 & = \lambda_{0r}^2 \begin{bmatrix} \mathbf{U}_0 \otimes \mathbf{W}_0 & \mathbf{U}_{0\perp} \otimes \mathbf{W}_0 \end{bmatrix} \begin{bmatrix} \mathbf{I}_r \otimes \begin{bmatrix} r \times r & \\ & \mathbf{I}_{p-r} \end{bmatrix} & \\ & \mathbf{I}_{p-r} \otimes \begin{bmatrix} \mathbf{I}_r & \\ & (p-r) \times (p-r) \end{bmatrix} \end{bmatrix} \\
1036 & \times \begin{bmatrix} \mathbf{U}_0^T \otimes \mathbf{W}_0^T \\ \mathbf{U}_{0\perp}^T \otimes \mathbf{W}_0^T \end{bmatrix} \\
1037 & = \lambda_{0r}^2 (\mathbf{U}_0 \otimes \mathbf{W}_0) \left\{ \mathbf{I}_r \otimes \begin{bmatrix} \mathbf{I}_r & \\ & \mathbf{I}_{p-r} \end{bmatrix} \right\} (\mathbf{U}_0^T \otimes \mathbf{W}_0^T) \\
1038 & + \lambda_{0r}^2 (\mathbf{U}_{0\perp} \otimes \mathbf{W}_0) \left\{ \mathbf{I}_{p-r} \otimes \begin{bmatrix} \mathbf{I}_r & \\ & \end{bmatrix} \right\} (\mathbf{U}_{0\perp}^T \otimes \mathbf{W}_0^T) \\
1039 & = \lambda_{0r}^2 (\mathbf{U}_0 \mathbf{U}_0^T \otimes \mathbf{U}_{0\perp} \mathbf{U}_{0\perp}^T + \mathbf{U}_{0\perp} \mathbf{U}_{0\perp}^T \otimes \mathbf{U}_0 \mathbf{U}_0^T) \\
1040 & = \lambda_{0r}^2 \{ \mathbf{U}_0 \mathbf{U}_0^T \otimes (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) + (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) \otimes \mathbf{U}_0 \mathbf{U}_0^T \}. \\
1041 &
\end{aligned}$$

1042 For the matrix $\Sigma_0^2 \otimes (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) + (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) \otimes \Sigma_0^2$, we write

$$\begin{aligned}
1043 & \Sigma_0^2 \otimes (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) + (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) \otimes \Sigma_0^2 \\
1044 &= \mathbf{W}_0 \widetilde{\mathbf{M}}_0 \mathbf{W}_0^T \otimes \mathbf{W}_0 \begin{bmatrix} r & \\ & \mathbf{I}_{p-r} \end{bmatrix} \mathbf{W}_0^T + \mathbf{W}_0 \begin{bmatrix} r & \\ & \mathbf{I}_{p-r} \end{bmatrix} \mathbf{W}_0^T \otimes \mathbf{W}_0 \widetilde{\mathbf{M}}_0 \mathbf{W}_0^T \\
1045 &= (\mathbf{W}_0 \otimes \mathbf{W}_0)(\widetilde{\mathbf{V}}_0 \otimes \widetilde{\mathbf{V}}_0) \left\{ \widetilde{\Lambda}_0^2 \otimes \begin{bmatrix} r & \\ & \mathbf{I}_{p-r} \end{bmatrix} + \begin{bmatrix} r & \\ & \mathbf{I}_{p-r} \end{bmatrix} \otimes \widetilde{\Lambda}_0^2 \right\} \\
1046 &\quad \times (\widetilde{\mathbf{V}}_0 \otimes \widetilde{\mathbf{V}}_0)^T (\mathbf{W}_0^T \otimes \mathbf{W}_0^T) \\
1047 &= (\mathbf{W}_0 \otimes \mathbf{W}_0)(\widetilde{\mathbf{V}}_0 \otimes \widetilde{\mathbf{V}}_0) \left[\begin{array}{c} \Lambda_0^2 \otimes \begin{bmatrix} r & \\ & \mathbf{I}_{p-r} \end{bmatrix} \\ \mathbf{I}_{p-r} \otimes \begin{bmatrix} \Lambda_0^2 & \\ & \end{bmatrix}_{(p-r) \times (p-r)} \end{array} \right] \\
1048 &\quad \times (\widetilde{\mathbf{V}}_0 \otimes \widetilde{\mathbf{V}}_0)^T (\mathbf{W}_0^T \otimes \mathbf{W}_0^T) \\
1049 &\succeq \lambda_{0r}^2 (\mathbf{W}_0 \otimes \mathbf{W}_0)(\widetilde{\mathbf{V}}_0 \otimes \widetilde{\mathbf{V}}_0) \left[\begin{array}{c} \mathbf{I}_r \otimes \begin{bmatrix} r & \\ & \mathbf{I}_{p-r} \end{bmatrix} \\ \mathbf{I}_{p-r} \otimes \begin{bmatrix} \mathbf{I}_r & \\ & \end{bmatrix}_{(p-r) \times (p-r)} \end{array} \right] \\
1050 &= \lambda_{0r}^2 (\mathbf{W}_0 \otimes \mathbf{W}_0) \left[\begin{array}{c} \mathbf{I}_r \otimes \begin{bmatrix} r \times r & \\ & \mathbf{I}_{p-r} \end{bmatrix} \\ \mathbf{I}_{p-r} \otimes \begin{bmatrix} \mathbf{I}_r & \\ & \end{bmatrix}_{(p-r) \times (p-r)} \end{array} \right] (\mathbf{W}_0^T \otimes \mathbf{W}_0^T) \\
1051 &= \lambda_{0r}^2 [\mathbf{U}_0 \otimes \mathbf{W}_0 \quad \mathbf{U}_{0\perp} \otimes \mathbf{W}_0] \left[\begin{array}{c} \mathbf{I}_r \otimes \begin{bmatrix} r \times r & \\ & \mathbf{I}_{p-r} \end{bmatrix} \\ \mathbf{I}_{p-r} \otimes \begin{bmatrix} \mathbf{I}_r & \\ & \end{bmatrix}_{(p-r) \times (p-r)} \end{array} \right] \\
1052 &\quad \times \begin{bmatrix} \mathbf{U}_0^T \otimes \mathbf{W}_0^T \\ \mathbf{U}_{0\perp}^T \otimes \mathbf{W}_0^T \end{bmatrix} \\
1053 &= \lambda_{0r}^2 (\mathbf{U}_0 \otimes \mathbf{W}_0) \left\{ \mathbf{I}_r \otimes \begin{bmatrix} & \\ & \mathbf{I}_{p-r} \end{bmatrix} \right\} (\mathbf{U}_0^T \otimes \mathbf{W}_0^T) \\
1054 &\quad + \lambda_{0r}^2 (\mathbf{U}_{0\perp} \otimes \mathbf{W}_0) \left\{ \mathbf{I}_{p-r} \otimes \begin{bmatrix} \mathbf{I}_r & \\ & \end{bmatrix} \right\} (\mathbf{U}_{0\perp}^T \otimes \mathbf{W}_0^T) \\
1055 &= \lambda_{0r}^2 (\mathbf{U}_0 \mathbf{U}_0^T \otimes \mathbf{U}_{0\perp} \mathbf{U}_{0\perp}^T) + \lambda_{0r}^2 (\mathbf{U}_{0\perp} \mathbf{U}_{0\perp}^T \otimes \mathbf{U}_0 \mathbf{U}_0^T) \\
1056 &= \lambda_{0r}^2 \{ \mathbf{U}_0 \mathbf{U}_0^T \otimes (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) + (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) \otimes \mathbf{U}_0 \mathbf{U}_0^T \},
\end{aligned}$$

1058 and the proof is thus completed. \square

1059 *Proof of Lemma 3.8.* Let $\mathbf{U}_{0\perp}$ be the orthogonal complement of \mathbf{U}_0 such that
1060 $[\mathbf{U}_0, \mathbf{U}_{0\perp}] \in \mathcal{O}(p)$, and one can therefore take $\mathbf{W}_0 = \mathbf{C}_0^{-T} \mathbf{C}_0$. For any $\mathbf{A} \in \mathbb{R}^{(p-r) \times r}$
1061 and $\boldsymbol{\varphi} = \text{vec}(\mathbf{A})$,

$$\begin{aligned}
1062 & \text{vec}(\mathbf{X}_{\boldsymbol{\varphi}})^T (\mathbf{C}_0^T \otimes \mathbf{C}_0^T) \{ \mathbf{U}_0 \mathbf{U}_0^T \otimes (\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) \} (\mathbf{C}_0 \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_{\boldsymbol{\varphi}}) \\
1063 &= \|(\mathbf{U}_0^T \otimes \mathbf{U}_{0\perp}^T)(\mathbf{C}_0 \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_{\boldsymbol{\varphi}})\|_2^2 \\
1064 &= \left\| \begin{bmatrix} r \times r & \mathbf{I}_{p-r} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{11}^T & \mathbf{C}_{21}^T \\ \mathbf{C}_{12}^T & \mathbf{C}_{22}^T \end{bmatrix} \begin{bmatrix} r \times r & -\mathbf{A}^T \\ \mathbf{A} & \end{bmatrix}_{(p-r) \times (p-r)} \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_r \\ \end{bmatrix}_{(p-r) \times r} \right\|_{\text{F}}^2 \\
1065 &= \|\mathbf{C}_{22}^T \mathbf{A} \mathbf{C}_{11} - \mathbf{C}_{12}^T \mathbf{A}^T \mathbf{C}_{21}\|_{\text{F}}^2,
\end{aligned}$$

and similarly,

$$\begin{aligned}
 & \text{vec}(\mathbf{X}_\varphi)^T (\mathbf{C}_0^T \otimes \mathbf{C}_0^T) \{(\mathbf{I}_p - \mathbf{U}_0 \mathbf{U}_0^T) \otimes \mathbf{U}_0 \mathbf{U}_0^T\} (\mathbf{C}_0 \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_\varphi) \\
 &= \|(\mathbf{U}_{0\perp}^T \otimes \mathbf{U}_0^T)(\mathbf{C}_0 \otimes \mathbf{C}_0) \text{vec}(\mathbf{X}_\varphi)\|_2^2 \\
 &= \left\| \begin{bmatrix} \mathbf{I}_r & r \times (p-r) \end{bmatrix} \begin{bmatrix} \mathbf{C}_{11}^T & \mathbf{C}_{21}^T \\ \mathbf{C}_{12}^T & \mathbf{C}_{22}^T \end{bmatrix} \begin{bmatrix} r \times r & -\mathbf{A}^T \\ \mathbf{A} & (p-r) \times (p-r) \end{bmatrix} \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} \begin{bmatrix} r \times r \\ \mathbf{I}_{p-r} \end{bmatrix} \right\|_F^2 \\
 &= \|\mathbf{C}_{22}^T \mathbf{A} \mathbf{C}_{11} - \mathbf{C}_{12}^T \mathbf{A}^T \mathbf{C}_{21}\|_F^2,
 \end{aligned}$$

and the proof of assertion (i) is completed. In the rest of the proof, we focus on assertion (ii), which is slightly involved. We consider two scenarios separately, i.e., $p \geq 2r$ and $p < 2r$.

Case I: $p - r \geq r$. Let $\mathbf{A}_0 = \mathbf{U}_{\mathbf{A}_0} \mathbf{S}_{\mathbf{A}_0} \mathbf{V}_{\mathbf{A}_0}^T$ be the singular value decomposition of \mathbf{A}_0 , where $\mathbf{U}_{\mathbf{A}_0} \in \mathbb{O}(p - r, r)$, $\mathbf{V}_{\mathbf{A}_0} \in \mathbb{O}(r)$, and $\mathbf{S}_{\mathbf{A}_0} = \text{diag}\{\sigma_1(\mathbf{A}_0), \dots, \sigma_r(\mathbf{A}_0)\}$. Suppose $\mathbf{U}_{\mathbf{A}_0\perp} \in \mathbb{O}(p - r, p - 2r)$ spans the orthogonal complement of $\text{Span}(\mathbf{U}_{\mathbf{A}_0})$, i.e., $\mathbf{U}_{\mathbf{A}_0}^T \mathbf{U}_{\mathbf{A}_0\perp} = \mathbf{0}_{r \times (p-2r)}$. Note that by assumption, $0 < \sigma_r(\mathbf{A}_0) \leq \sigma_1(\mathbf{A}_0) < 1$. Using the property of commutation matrices [43] that

$$\begin{aligned}
 & (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T) \mathbf{K}_{(p-r)r} = \mathbf{K}_{r(p-r)} (\mathbf{C}_{12}^T \otimes \mathbf{C}_{21}^T), \\
 & \mathbf{K}_{r(p-r)} (\mathbf{U}_{\mathbf{A}_0} \otimes \mathbf{V}_{\mathbf{A}_0}) = (\mathbf{V}_{\mathbf{A}_0} \otimes \mathbf{U}_{\mathbf{A}_0}) \mathbf{K}_{rr},
 \end{aligned}$$

we write

$$\begin{aligned}
 & \mathbf{C}_{11} \otimes \mathbf{C}_{22} - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T) \mathbf{K}_{(p-r)r} \\
 &= \mathbf{C}_{11} \otimes \mathbf{C}_{22} - \mathbf{K}_{r(p-r)} (\mathbf{C}_{12}^T \otimes \mathbf{C}_{21}^T) \\
 &= \mathbf{V}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{V}_{\mathbf{A}_0}^T \otimes \{\mathbf{I}_{p-r} - \mathbf{U}_{\mathbf{A}_0} \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \mathbf{U}_{\mathbf{A}_0}^T\} \\
 &\quad + \mathbf{K}_{r(p-r)} \{\mathbf{U}_{\mathbf{A}_0} \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{V}_{\mathbf{A}_0}^T \otimes \mathbf{V}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \mathbf{U}_{\mathbf{A}_0}^T\} \\
 &= \mathbf{V}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{V}_{\mathbf{A}_0}^T \otimes (\mathbf{U}_{\mathbf{A}_0} \mathbf{U}_{\mathbf{A}_0}^T + \mathbf{U}_{\mathbf{A}_0\perp} \mathbf{U}_{\mathbf{A}_0\perp}^T) \\
 &\quad - \mathbf{V}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{V}_{\mathbf{A}_0}^T \otimes \{\mathbf{U}_{\mathbf{A}_0} \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \mathbf{U}_{\mathbf{A}_0}^T\} \\
 &\quad + \mathbf{K}_{r(p-r)} (\mathbf{U}_{\mathbf{A}_0} \otimes \mathbf{V}_{\mathbf{A}_0}) \{\mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \otimes (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0}\} (\mathbf{V}_{\mathbf{A}_0}^T \otimes \mathbf{U}_{\mathbf{A}_0}^T) \\
 &= (\mathbf{V}_{\mathbf{A}_0} \otimes \mathbf{U}_{\mathbf{A}_0}) [(\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \otimes \{\mathbf{I}_r - \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0}\}] (\mathbf{V}_{\mathbf{A}_0}^T \otimes \mathbf{U}_{\mathbf{A}_0}^T) \\
 &\quad + (\mathbf{V}_{\mathbf{A}_0} \otimes \mathbf{U}_{\mathbf{A}_0\perp}) \{(\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \otimes \mathbf{I}_{p-r}\} (\mathbf{V}_{\mathbf{A}_0}^T \otimes \mathbf{U}_{\mathbf{A}_0\perp}^T) \\
 &\quad + (\mathbf{V}_{\mathbf{A}_0} \otimes \mathbf{U}_{\mathbf{A}_0}) \mathbf{K}_{rr} \{(\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \otimes \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1}\} (\mathbf{V}_{\mathbf{A}_0}^T \otimes \mathbf{U}_{\mathbf{A}_0}^T) \\
 &= (\mathbf{V}_{\mathbf{A}_0} \otimes \mathbf{U}_{\mathbf{A}_0}) \tilde{\mathbf{R}} (\mathbf{V}_{\mathbf{A}_0}^T \otimes \mathbf{U}_{\mathbf{A}_0}^T) \\
 &\quad + (\mathbf{V}_{\mathbf{A}_0} \otimes \mathbf{U}_{\mathbf{A}_0\perp}) \{(\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \otimes \mathbf{I}_{p-r}\} (\mathbf{V}_{\mathbf{A}_0}^T \otimes \mathbf{U}_{\mathbf{A}_0\perp}^T),
 \end{aligned}$$

where

$$\begin{aligned}
 & \tilde{\mathbf{R}} := (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \otimes \{\mathbf{I}_r - \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0}\} \\
 &\quad + \mathbf{K}_{rr} \{(\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \otimes \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1}\} \\
 &= (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \otimes (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} - (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \otimes \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \\
 &\quad + (\mathbf{K}_{rr} + \mathbf{I}_{r^2}) \{(\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \otimes \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1}\}.
 \end{aligned}$$

Note that $[\mathbf{V}_{\mathbf{A}_0} \otimes \mathbf{U}_{\mathbf{A}_0}, \mathbf{V}_{\mathbf{A}_0} \otimes \mathbf{U}_{\mathbf{A}_0\perp}] \in \mathbb{O}(r(p - r))$. It follows that

$$\begin{aligned}
 & \sigma_{\min}\{\mathbf{C}_{11}^T \otimes \mathbf{C}_{22}^T - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T) \mathbf{K}_{(p-r)r}\} \\
 &= \min \left[\sigma_{\min}(\tilde{\mathbf{R}}), \sigma_{\min}\{(\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \otimes \mathbf{I}_{p-r}\} \right].
 \end{aligned}$$

1108 We now provide a lower bound for the smallest singular value of $\tilde{\mathbf{R}}$. When $r = 1$,
 1109 $\mathbf{K}_{rr} = \mathbf{K}_{11} = 1$, $\mathbf{S}_{\mathbf{A}_0} = \|\mathbf{A}_0\|_2 \in \mathbb{R}^{1 \times 1}$, and hence

$$1110 \quad \tilde{\mathbf{R}} = \frac{1}{(1 + \mathbf{S}_{\mathbf{A}_0}^2)^2} + \frac{\mathbf{S}_{\mathbf{A}_0}^2}{(1 + \mathbf{S}_{\mathbf{A}_0}^2)^2} = \frac{1}{1 + \|\mathbf{A}\|_2^2}.$$

1111 When $r > 1$, we consider the following approach. Denote the diagonal matrices

$$1112 \quad \mathbf{D}_1 = (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1}, \quad \mathbf{D}_2 = (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0}.$$

1113 Note that by the property of the commutation matrix,

$$1114 \quad (\mathbf{K}_{rr} + \mathbf{I}_{r^2}) \{ (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \otimes \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \} \\ 1115 \quad = (\mathbf{K}_{rr} + \mathbf{I}_{r^2}) (\mathbf{D}_2 \otimes \mathbf{D}_2) = (\mathbf{D}_2 \otimes \mathbf{D}_2) (\mathbf{K}_{rr} + \mathbf{I}_{r^2}).$$

1117 Therefore, $(\mathbf{K}_{rr} + \mathbf{I}_{r^2})(\mathbf{D}_2 \otimes \mathbf{D}_2)$ is symmetric and is a product two positive semidefinite
 1118 matrices $\mathbf{K}_{rr} + \mathbf{I}_{r^2}$ and $\mathbf{D}_2 \otimes \mathbf{D}_2$ because \mathbf{K}_{rr} is symmetric and only has eigenvalues
 1119 in $\{\pm 1\}$ [43]. Therefore, by Corollary 11 in [63], we have

$$1120 \quad \lambda_{\min} \{ (\mathbf{K}_{rr} + \mathbf{I}_{r^2})(\mathbf{D}_2 \otimes \mathbf{D}_2) \} \geq \lambda_{\min}(\mathbf{K}_{rr} + \mathbf{I}_{r^2}) \lambda_{\min}(\mathbf{D}_2 \otimes \mathbf{D}_2) = 0.$$

1121 This further implies that $(\mathbf{K}_{rr} + \mathbf{I}_{r^2})(\mathbf{D}_2 \otimes \mathbf{D}_2)$ is positive semidefinite and $\tilde{\mathbf{R}}$ is
 1122 symmetric. Note that $\mathbf{D}_1 \succeq \mathbf{D}_2$ because $\sigma_k(\mathbf{A}_0) \in [0, 1]$, and that $\tilde{\mathbf{R}}$ is positive
 1123 semidefinite. Hence, to provide a lower bound for the smallest singular value of $\tilde{\mathbf{R}}$, it
 1124 is sufficient to provide a lower bound for the smallest eigenvalue of

$$1125 \quad (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \otimes (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} - (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \otimes \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \\ 1126 \quad = \mathbf{D}_1 \otimes \mathbf{D}_1 - \mathbf{D}_2 \otimes \mathbf{D}_2.$$

1128 Write

$$1129 \quad \mathbf{D}_1 \otimes \mathbf{D}_1 - \mathbf{D}_2 \otimes \mathbf{D}_2 \\ 1130 \quad = \text{diag} \left\{ \frac{1}{1 + \sigma_1^2(\mathbf{A}_0)}, \dots, \frac{1}{1 + \sigma_r^2(\mathbf{A}_0)} \right\} \otimes \text{diag} \left\{ \frac{1}{1 + \sigma_1^2(\mathbf{A}_0)}, \dots, \frac{1}{1 + \sigma_r^2(\mathbf{A}_0)} \right\} \\ 1131 \quad - \text{diag} \left\{ \frac{\sigma_1(\mathbf{A}_0)}{1 + \sigma_1^2(\mathbf{A}_0)}, \dots, \frac{\sigma_r(\mathbf{A}_0)}{1 + \sigma_r^2(\mathbf{A}_0)} \right\} \otimes \text{diag} \left\{ \frac{\sigma_1(\mathbf{A}_0)}{1 + \sigma_1^2(\mathbf{A}_0)}, \dots, \frac{\sigma_r(\mathbf{A}_0)}{1 + \sigma_r^2(\mathbf{A}_0)} \right\} \\ 1132 \quad = \begin{bmatrix} \frac{1}{1 + \sigma_1^2(\mathbf{A}_0)} \mathbf{D}_1 - \frac{\sigma_1(\mathbf{A}_0)}{1 + \sigma_1^2(\mathbf{A}_0)} \mathbf{D}_2 & & \\ & \ddots & \\ & & \frac{1}{1 + \sigma_r^2(\mathbf{A}_0)} \mathbf{D}_1 - \frac{\sigma_r(\mathbf{A}_0)}{1 + \sigma_r^2(\mathbf{A}_0)} \mathbf{D}_2 \end{bmatrix} \succeq_{r^2 \times r^2}.$$

1134 It follows that

$$1135 \quad \sigma_{\min}(\tilde{\mathbf{R}}) = \sigma_{\min} \{ \mathbf{D}_1 \otimes \mathbf{D}_1 - \mathbf{D}_2 \otimes \mathbf{D}_2 + (\mathbf{K}_{rr} + \mathbf{I}_{r^2})(\mathbf{D}_2 \otimes \mathbf{D}_2) \} \\ 1136 \quad = \lambda_{\min} \{ \mathbf{D}_1 \otimes \mathbf{D}_1 - \mathbf{D}_2 \otimes \mathbf{D}_2 + (\mathbf{K}_{rr} + \mathbf{I}_{r^2})(\mathbf{D}_2 \otimes \mathbf{D}_2) \} \\ 1137 \quad \geq \lambda_{\min} \{ \mathbf{D}_1 \otimes \mathbf{D}_1 - \mathbf{D}_2 \otimes \mathbf{D}_2 \} \\ 1138 \quad = \min_{k \in [r]} \lambda_{\min} \left\{ \frac{1}{1 + \sigma_k^2(\mathbf{A}_0)} \mathbf{D}_1 - \frac{\sigma_k(\mathbf{A}_0)}{1 + \sigma_k^2(\mathbf{A}_0)} \mathbf{D}_2 \right\} \\ 1139 \quad = \min_{k, l \in [r]} \frac{1 - \sigma_k(\mathbf{A}_0) \sigma_l(\mathbf{A}_0)}{\{1 + \sigma_k^2(\mathbf{A}_0)\} \{1 + \sigma_l^2(\mathbf{A}_0)\}}. \\ 1140$$

1141 We finally conclude that when $r > 1$,

$$\begin{aligned}
 1142 \quad & \sigma_{\min}\{\mathbf{C}_{11} \otimes \mathbf{C}_{22} - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T)\mathbf{K}_{(p-r)r}\} \\
 1143 \quad & = \lambda_{\min}\{\mathbf{C}_{11} \otimes \mathbf{C}_{22} - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T)\mathbf{K}_{(p-r)r}\} \\
 1144 \quad & = \min\left\{\sigma_{\min}(\tilde{\mathbf{R}}), \sigma_{\min}(\mathbf{D}_1 \otimes \mathbf{I}_{p-r})\right\} \\
 1145 \quad & \geq \min\left[\min_{k,l \in [r]} \frac{1 - \sigma_k(\mathbf{A}_0)\sigma_l(\mathbf{A}_0)}{\{1 + \sigma_k^2(\mathbf{A}_0)\}\{1 + \sigma_l^2(\mathbf{A}_0)\}}, \min_{k \in [r]} \frac{1}{1 + \sigma_k^2(\mathbf{A}_0)}\right] \\
 1146 \quad & = \frac{1 - \|\mathbf{A}_0\|_2^2}{(1 + \|\mathbf{A}_0\|_2^2)^2}, \\
 1147 \quad &
 \end{aligned}$$

1148 and when $r = 1$, we directly obtain

$$\begin{aligned}
 1149 \quad & \sigma_{\min}\{\mathbf{C}_{11} \otimes \mathbf{C}_{22} - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T)\mathbf{K}_{(p-r)r}\} \\
 1150 \quad & = \lambda_{\min}\{\mathbf{C}_{11} \otimes \mathbf{C}_{22} - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T)\mathbf{K}_{(p-r)r}\} \\
 1151 \quad & = \min\left\{\sigma_{\min}(\tilde{\mathbf{R}}), \sigma_{\min}(\mathbf{D}_1 \otimes \mathbf{I}_{p-r})\right\} \\
 1152 \quad & \geq \min\left[\min_{k,l \in [r]} \frac{1 - \sigma_k(\mathbf{A}_0)\sigma_l(\mathbf{A}_0)}{\{1 + \sigma_k^2(\mathbf{A}_0)\}\{1 + \sigma_l^2(\mathbf{A}_0)\}}, \min_{k \in [r]} \frac{1}{1 + \sigma_k^2(\mathbf{A}_0)}\right] \\
 1153 \quad & = \frac{1}{1 + \|\mathbf{A}_0\|_2^2}. \\
 1154 \quad &
 \end{aligned}$$

1155 **Case II:** $p - r < r$. This situation occurs only if $r > 1$. Let $\mathbf{A}_0 = \mathbf{U}_{\mathbf{A}_0} \mathbf{S}_{\mathbf{A}_0} \mathbf{V}_{\mathbf{A}_0}^T$ be
 1156 the singular value decomposition of \mathbf{A}_0 , where $\mathbf{U}_{\mathbf{A}_0} \in \mathbb{O}(p - r)$, $\mathbf{V}_{\mathbf{A}_0} \in \mathbb{O}(r, p - r)$,
 1157 and $\mathbf{S}_{\mathbf{A}_0} = \text{diag}\{\sigma_1(\mathbf{A}_0), \dots, \sigma_{p-r}(\mathbf{A}_0)\}$. Suppose $\mathbf{V}_{\mathbf{A}_0\perp} \in \mathbb{O}(r, 2r - p)$ spans the
 1158 orthogonal complement of $\text{Span}(\mathbf{V}_{\mathbf{A}_0})$, i.e., $\mathbf{V}_{\mathbf{A}_0}^T \mathbf{V}_{\mathbf{A}_0\perp} = \mathbf{0}_{p-r \times (2r-p)}$. Note that by
 1159 assumption, $0 < \sigma_r(\mathbf{A}_0) \leq \sigma_1(\mathbf{A}_0) < 1$. Observe that

$$\begin{aligned}
 1160 \quad & \mathbf{C}_{11} = (\mathbf{I}_r + \mathbf{A}_0^T \mathbf{A}_0)^{-1} = [\mathbf{V}_{\mathbf{A}_0} \quad \mathbf{V}_{\mathbf{A}_0\perp}] \begin{bmatrix} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} & \\ & \mathbf{I}_{2r-p} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{\mathbf{A}_0}^T \\ \mathbf{V}_{\mathbf{A}_0\perp}^T \end{bmatrix}, \\
 1161 \quad &
 \end{aligned}$$

1162 and

$$\begin{aligned}
 1163 \quad & \mathbf{C}_{22} = \mathbf{I}_{p-r} - \mathbf{A}_0(\mathbf{I}_r + \mathbf{A}_0^T \mathbf{A}_0)^{-1} \mathbf{A}_0^T \\
 1164 \quad & = \mathbf{I}_{p-r} - \mathbf{U}_{\mathbf{A}_0} \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \mathbf{U}_{\mathbf{A}_0}^T = \mathbf{U}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{U}_{\mathbf{A}_0}^T.
 \end{aligned}$$

1166 Also,

$$\begin{aligned}
 1167 \quad & \mathbf{C}_{12}^T = -\mathbf{A}_0(\mathbf{I}_r + \mathbf{A}_0^T \mathbf{A}_0)^{-1} = -\mathbf{U}_{\mathbf{A}_0} \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{V}_{\mathbf{A}_0}^T, \\
 1168 \quad & \mathbf{C}_{21}^T = (\mathbf{I}_r + \mathbf{A}_0^T \mathbf{A}_0)^{-1} \mathbf{A}_0^T = \mathbf{V}_{\mathbf{A}_0} (\mathbf{I}_r + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \mathbf{U}_{\mathbf{A}_0}^T.
 \end{aligned}$$

1170 Using the property of commutation matrices [43] that

$$\begin{aligned}
 1171 \quad & (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T) \mathbf{K}_{(p-r)r} = \mathbf{K}_{r(p-r)} (\mathbf{C}_{12}^T \otimes \mathbf{C}_{21}^T), \\
 1172 \quad & \mathbf{K}_{r(p-r)} (\mathbf{U}_{\mathbf{A}_0} \otimes \mathbf{V}_{\mathbf{A}_0}) = (\mathbf{V}_{\mathbf{A}_0} \otimes \mathbf{U}_{\mathbf{A}_0}) \mathbf{K}_{(p-r)(p-r)},
 \end{aligned}$$

we write

$$\begin{aligned}
& \mathbf{C}_{11} \otimes \mathbf{C}_{22} - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T) \mathbf{K}_{(p-r)r} \\
&= \mathbf{C}_{11} \otimes \mathbf{C}_{22} - \mathbf{K}_{r(p-r)} (\mathbf{C}_{12}^T \otimes \mathbf{C}_{21}^T) \\
&= [\mathbf{V}_{\mathbf{A}_0} \quad \mathbf{V}_{\mathbf{A}_0 \perp}] \begin{bmatrix} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} & \\ & \mathbf{I}_{2r-p} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{\mathbf{A}_0}^T \\ \mathbf{V}_{\mathbf{A}_0 \perp}^T \end{bmatrix} \otimes \mathbf{U}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{U}_{\mathbf{A}_0}^T \\
&\quad + \mathbf{K}_{r(p-r)} \{ \mathbf{U}_{\mathbf{A}_0} \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{V}_{\mathbf{A}_0}^T \otimes \mathbf{V}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \mathbf{U}_{\mathbf{A}_0}^T \} \\
&= \{ \mathbf{V}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{V}_{\mathbf{A}_0}^T + \mathbf{V}_{\mathbf{A}_0 \perp} \mathbf{V}_{\mathbf{A}_0 \perp}^T \} \otimes \{ \mathbf{U}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{U}_{\mathbf{A}_0}^T \} \\
&\quad + \mathbf{K}_{r(p-r)} (\mathbf{U}_{\mathbf{A}_0} \otimes \mathbf{V}_{\mathbf{A}_0}) \{ \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \otimes (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \} \\
&\quad \times (\mathbf{V}_{\mathbf{A}_0}^T \otimes \mathbf{U}_{\mathbf{A}_0}^T) \\
&= (\mathbf{V}_{\mathbf{A}_0} \otimes \mathbf{U}_{\mathbf{A}_0}) \{ (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \otimes (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \} (\mathbf{V}_{\mathbf{A}_0}^T \otimes \mathbf{U}_{\mathbf{A}_0}^T) \\
&\quad + (\mathbf{V}_{\mathbf{A}_0 \perp} \otimes \mathbf{U}_{\mathbf{A}_0}) \{ \mathbf{I}_{2r-p} \otimes (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \} (\mathbf{V}_{\mathbf{A}_0 \perp}^T \otimes \mathbf{U}_{\mathbf{A}_0}^T) \\
&\quad + (\mathbf{V}_{\mathbf{A}_0} \otimes \mathbf{U}_{\mathbf{A}_0}) \mathbf{K}_{(p-r)(p-r)} \{ (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \otimes \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \} \\
&\quad \times (\mathbf{V}_{\mathbf{A}_0}^T \otimes \mathbf{U}_{\mathbf{A}_0}^T) \\
&= (\mathbf{V}_{\mathbf{A}_0} \otimes \mathbf{U}_{\mathbf{A}_0}) \tilde{\mathbf{R}} (\mathbf{V}_{\mathbf{A}_0}^T \otimes \mathbf{U}_{\mathbf{A}_0}^T) \\
&\quad + (\mathbf{V}_{\mathbf{A}_0 \perp} \otimes \mathbf{U}_{\mathbf{A}_0}) \{ \mathbf{I}_{2r-p} \otimes (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \} (\mathbf{V}_{\mathbf{A}_0 \perp}^T \otimes \mathbf{U}_{\mathbf{A}_0}^T),
\end{aligned}$$

where

$$\begin{aligned}
& \tilde{\mathbf{R}} = (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \otimes (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \\
&\quad + \mathbf{K}_{(p-r)(p-r)} \{ (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \otimes \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \} \\
&= (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \otimes (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} - (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \otimes \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \\
&\quad + (\mathbf{K}_{(p-r)(p-r)} + \mathbf{I}_{(p-r)^2}) \{ (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \otimes \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \}.
\end{aligned}$$

Note that $[\mathbf{V}_{\mathbf{A}_0} \otimes \mathbf{U}_{\mathbf{A}_0}, \mathbf{V}_{\mathbf{A}_0 \perp} \otimes \mathbf{U}_{\mathbf{A}_0}] \in \mathbb{O}(r(p-r))$. It follows that

$$\sigma_{\min} \{ \mathbf{C}_{11} \otimes \mathbf{C}_{22} - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T) \mathbf{K}_{(p-r)r} \} = \sigma_{\min}(\tilde{\mathbf{R}}) \wedge \sigma_{\min} \{ \mathbf{I}_{2r-p} \otimes (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \}.$$

Similar to the case where $p-r \geq r$, we also provide a lower bound for the smallest singular value of $\tilde{\mathbf{R}}$. Denote the diagonal matrices

$$\mathbf{D}_1 = (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1}, \quad \mathbf{D}_2 = (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0}.$$

Note that by the property of the commutation matrix,

$$\begin{aligned}
& (\mathbf{K}_{(p-r)(p-r)} + \mathbf{I}_{(p-r)^2}) \{ (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \otimes \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \} \\
&= (\mathbf{K}_{(p-r)(p-r)} + \mathbf{I}_{(p-r)^2}) (\mathbf{D}_2 \otimes \mathbf{D}_2) = (\mathbf{D}_2 \otimes \mathbf{D}_2) (\mathbf{K}_{(p-r)(p-r)} + \mathbf{I}_{(p-r)^2}).
\end{aligned}$$

Therefore, $(\mathbf{K}_{(p-r)(p-r)} + \mathbf{I}_{(p-r)^2}) (\mathbf{D}_2 \otimes \mathbf{D}_2)$ is symmetric and is a product two positive semidefinite matrices $\mathbf{K}_{(p-r)(p-r)} + \mathbf{I}_{(p-r)^2}$ and $\mathbf{D}_2 \otimes \mathbf{D}_2$. Again, by Corollary 11 in [63], we have

$$\begin{aligned}
& \lambda_{\min} \{ (\mathbf{K}_{(p-r)(p-r)} + \mathbf{I}_{(p-r)^2}) (\mathbf{D}_2 \otimes \mathbf{D}_2) \} \\
& \geq \lambda_{\min} (\mathbf{K}_{(p-r)(p-r)} + \mathbf{I}_{(p-r)^2}) \lambda_{\min} (\mathbf{D}_2 \otimes \mathbf{D}_2) = 0.
\end{aligned}$$

This further implies that $(\mathbf{K}_{(p-r)(p-r)} + \mathbf{I}_{(p-r)^2}) (\mathbf{D}_2 \otimes \mathbf{D}_2)$ is positive semidefinite and $\tilde{\mathbf{R}}$ is symmetric. Note that $\mathbf{D}_1 \succeq \mathbf{D}_2$ because $\sigma_k(\mathbf{A}_0) \in [0, 1]$, and that $\tilde{\mathbf{R}}$ is positive

semidefinite. Hence, to provide a lower bound for the smallest singular value of $\tilde{\mathbf{R}}$, it is sufficient to provide a lower bound for the smallest eigenvalue of

$$(\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \otimes (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} - (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \mathbf{S}_{\mathbf{A}_0} \otimes \mathbf{S}_{\mathbf{A}_0} (\mathbf{I}_{p-r} + \mathbf{S}_{\mathbf{A}_0}^2)^{-1} \\ = \mathbf{D}_1 \otimes \mathbf{D}_1 - \mathbf{D}_2 \otimes \mathbf{D}_2.$$

Write

$$\mathbf{D}_1 \otimes \mathbf{D}_1 - \mathbf{D}_2 \otimes \mathbf{D}_2 \\ = \text{diag} \left\{ \frac{1}{1 + \sigma_1^2(\mathbf{A}_0)}, \dots, \frac{1}{1 + \sigma_{p-r}^2(\mathbf{A}_0)} \right\} \\ \otimes \text{diag} \left\{ \frac{1}{1 + \sigma_1^2(\mathbf{A}_0)}, \dots, \frac{1}{1 + \sigma_{p-r}^2(\mathbf{A}_0)} \right\} \\ - \text{diag} \left\{ \frac{\sigma_1(\mathbf{A}_0)}{1 + \sigma_1^2(\mathbf{A}_0)}, \dots, \frac{\sigma_{p-r}(\mathbf{A}_0)}{1 + \sigma_{p-r}^2(\mathbf{A}_0)} \right\} \\ \otimes \text{diag} \left\{ \frac{\sigma_1(\mathbf{A}_0)}{1 + \sigma_1^2(\mathbf{A}_0)}, \dots, \frac{\sigma_{p-r}(\mathbf{A}_0)}{1 + \sigma_{p-r}^2(\mathbf{A}_0)} \right\} \\ = \begin{bmatrix} \frac{1}{1 + \sigma_1^2(\mathbf{A}_0)} \mathbf{D}_1 - \frac{\sigma_1(\mathbf{A}_0)}{1 + \sigma_1^2(\mathbf{A}_0)} \mathbf{D}_2 & & \\ & \ddots & \\ & & \frac{1}{1 + \sigma_{p-r}^2(\mathbf{A}_0)} \mathbf{D}_1 - \frac{\sigma_{p-r}(\mathbf{A}_0)}{1 + \sigma_{p-r}^2(\mathbf{A}_0)} \mathbf{D}_2 \end{bmatrix} \\ \succeq_{p-r^2 \times p-rb^2}.$$

It follows that

$$\sigma_{\min}(\tilde{\mathbf{R}}) = \sigma_{\min} \{ \mathbf{D}_1 \otimes \mathbf{D}_1 - \mathbf{D}_2 \otimes \mathbf{D}_2 + (\mathbf{K}_{(p-r)(p-r)} + \mathbf{I}_{(p-r)^2})(\mathbf{D}_2 \otimes \mathbf{D}_2) \} \\ = \lambda_{\min} \{ \mathbf{D}_1 \otimes \mathbf{D}_1 - \mathbf{D}_2 \otimes \mathbf{D}_2 + (\mathbf{K}_{(p-r)(p-r)} + \mathbf{I}_{(p-r)^2})(\mathbf{D}_2 \otimes \mathbf{D}_2) \} \\ \geq \lambda_{\min} \{ \mathbf{D}_1 \otimes \mathbf{D}_1 - \mathbf{D}_2 \otimes \mathbf{D}_2 \} \\ = \min_{k \in [p-r]} \lambda_{\min} \left\{ \frac{1}{1 + \sigma_k^2(\mathbf{A}_0)} \mathbf{D}_1 - \frac{\sigma_k(\mathbf{A}_0)}{1 + \sigma_k^2(\mathbf{A}_0)} \mathbf{D}_2 \right\} \\ = \min_{k, l \in [p-r]} \frac{1 - \sigma_k(\mathbf{A}_0) \sigma_l(\mathbf{A}_0)}{\{1 + \sigma_k^2(\mathbf{A}_0)\} \{1 + \sigma_l^2(\mathbf{A}_0)\}}.$$

We finally conclude that

$$\sigma_{\min} \{ \mathbf{C}_{11} \otimes \mathbf{C}_{22} - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T) \mathbf{K}_{(p-r)r} \} \\ = \lambda_{\min} \{ \mathbf{C}_{11} \otimes \mathbf{C}_{22} - (\mathbf{C}_{21}^T \otimes \mathbf{C}_{12}^T) \mathbf{K}_{(p-r)r} \} \\ = \min \left\{ \sigma_{\min}(\tilde{\mathbf{R}}), \sigma_{\min}(\mathbf{I}_{2r-p} \otimes \mathbf{D}_1) \right\} \\ \geq \min \left[\min_{k, l \in [p-r]} \frac{1 - \sigma_k(\mathbf{A}_0) \sigma_l(\mathbf{A}_0)}{\{1 + \sigma_k^2(\mathbf{A}_0)\} \{1 + \sigma_l^2(\mathbf{A}_0)\}}, \min_{k \in [p-r]} \frac{1}{1 + \sigma_k^2(\mathbf{A}_0)} \right] \\ = \frac{1 - \|\mathbf{A}_0\|_2^2}{(1 + \|\mathbf{A}_0\|_2^2)^2}.$$

The proof is thus completed. \square

Joint proof of Propositions 3.9 and 4.1. By Example 4.2 in [22], it is sufficient to show that $\mathcal{M}(p, r)$ and $\mathcal{G}(p, r)$ are regular surfaces in $\mathbb{R}^{p \times p}$ in the following sense:

For every $\Sigma \in \mathcal{M}(p, r)$ and $\Pi \in \mathcal{G}(p, r)$, there exist neighborhoods \mathcal{V}_1 of Σ and \mathcal{V}_2 of Π in $\mathbb{R}^{p \times p}$ and permutation matrices $\mathbf{P}_1, \mathbf{P}_2 \in \mathcal{P}(p)$, such that the maps $\mathbf{x}_{\mathbf{P}_1} : \mathcal{U} := \mathcal{D}(p, r) \rightarrow \mathcal{M}(p, r) \cap \mathcal{V}$ onto $\mathcal{M}(p, r) \cap \mathcal{V}_1$ and $\mathbf{y}_{\mathbf{P}_2} : \mathcal{A} \rightarrow \mathcal{G}(p, r) \cap \mathcal{V}_2$ onto $\mathcal{G}(p, r) \cap \mathcal{V}_2$ satisfy

- (a) $\mathbf{x}_{\mathbf{P}_1}$ and $\mathbf{y}_{\mathbf{P}_2}$ are differentiable and are homeomorphisms;
- (b) The Jacobians $\partial \text{vec}\{\mathbf{x}_{\mathbf{P}_1}(\boldsymbol{\theta})\} / \partial \boldsymbol{\theta}^T$ and $\partial \text{vec}\{\mathbf{y}_{\mathbf{P}_2}(\boldsymbol{\varphi})\} / \partial \boldsymbol{\varphi}^T$ have ranks d and $(p-r)r$, respectively.

Let $\Sigma_0 \in \mathcal{M}(p, r)$ and suppose it has eigendecomposition $\Sigma_0 = \mathbf{U}_\Sigma \mathbf{S}_\Sigma \mathbf{U}_\Sigma^T$, where \mathbf{S}_Σ is the diagonal matrix of the nonzero eigenvalues of Σ and $\mathbf{U}_\Sigma \in \mathbb{O}(p, r)$. Similarly, let $\Pi_0 \in \mathcal{G}(p, r)$ with $\Pi_0 = \mathbf{Q}\mathbf{Q}^T$ for some $\mathbf{Q} \in \mathbb{O}(p, r)$. Clearly, there exist permutation matrices $\mathbf{P}_1, \mathbf{P}_2 \in \{0, 1\}^{p \times p}$ and $\mathbf{V}_1, \mathbf{V}_2 \in \mathbb{O}(r)$ such that $\mathbf{P}_1^T \mathbf{U}_\Sigma \mathbf{V}_1, \mathbf{P}_2^T \mathbf{Q} \mathbf{V}_2 \in \mathbb{O}_+(p, r)$. This implies that $\mathbf{P}^T \Sigma_0 \mathbf{P} \in \mathcal{S}(p, r)$ and hence, there exists $\mathbf{A}_{01}, \mathbf{A}_{02} \in \mathbb{R}^{(p-r) \times r}$ with $\|\mathbf{A}_{01}\|_2, \|\mathbf{A}_{02}\|_2 < 1$ and $\boldsymbol{\varphi}_{01} = \text{vec}(\mathbf{A}_{01}), \boldsymbol{\varphi}_{02} = \text{vec}(\mathbf{A}_{02}) \in \mathcal{A}$, $\boldsymbol{\theta}_0 = [\boldsymbol{\varphi}_{01}^T, \boldsymbol{\mu}_0^T]^T \in \mathcal{D}(p, r)$, such that $\Sigma_0 = \mathbf{P}_1 \Sigma(\boldsymbol{\theta}_0) \mathbf{P}_1^T$ and $\Pi_0 = \mathbf{P}_2 \mathbf{U}(\boldsymbol{\varphi}_{02}) \mathbf{U}(\boldsymbol{\varphi}_{02})^T \mathbf{P}_2^T$. We next verify that $\mathbf{x}_{\mathbf{P}_1}$ and $\mathbf{y}_{\mathbf{P}_2}$ have the desired properties.

- (i) Clearly, $\mathbf{x}_{\mathbf{P}_1}$ is differentiable since $\text{vec}\{\mathbf{x}_{\mathbf{P}_1}(\boldsymbol{\theta})\} = (\mathbf{P}_1^T \otimes \mathbf{P}_1^T) \text{vec}\{\Sigma(\boldsymbol{\theta})\}$ and $\Sigma(\cdot)$ is differentiable. Similarly, $\mathbf{y}_{\mathbf{P}_2}$ is also differentiable.
- (ii) We show that the image of $\mathbf{x}_{\mathbf{P}_1} : \mathcal{U} \rightarrow \mathcal{M}(p, r)$ is given by

$$\mathbf{P}_1 \mathcal{S}(p, r) \mathbf{P}_1^T := \{\mathbf{P}_1 \Sigma \mathbf{P}_1^T : \Sigma \in \mathcal{S}(p, r)\}$$

and the image of $\mathbf{y}_{\mathbf{P}_2} : \mathcal{A} \rightarrow \mathcal{G}(p, r)$ is given by

$$\{\mathbf{P}_2 \mathbf{U} \mathbf{U}^T \mathbf{P}_2^T : \mathbf{U} \in \mathbb{O}_+(p, r)\}.$$

Indeed, if $\Sigma' \in \mathbf{P}_1 \mathcal{S}(p, r) \mathbf{P}_1^T$, then there exists $\Sigma'' \in \mathcal{S}(p, r)$ such that $\Sigma' = \mathbf{P}_1 \Sigma'' \mathbf{P}_1^T$. It follows by the parameterization function $\Sigma(\cdot)$ defined in (2.6) that there exists some $\boldsymbol{\theta} \in \mathcal{D}(p, r)$ such that $\Sigma(\boldsymbol{\theta}) = \Sigma''$, and hence,

$$\mathbf{x}_{\mathbf{P}_1}(\boldsymbol{\theta}) = \mathbf{P}_1 \Sigma(\boldsymbol{\theta}) \mathbf{P}_1^T = \mathbf{P}_1 \Sigma'' \mathbf{P}_1^T = \Sigma'.$$

This shows that $\Sigma' \in \mathbf{x}_{\mathbf{P}_1}(\mathcal{U})$. Conversely, if $\Sigma' \in \mathbf{x}_{\mathbf{P}_1}(\mathcal{U})$, it is clear that $\mathbf{x}_{\mathbf{P}_1}(\boldsymbol{\theta}) = \mathbf{P}_1 \Sigma(\boldsymbol{\theta}) \mathbf{P}_1^T = \Sigma' \in \mathbf{P}_1 \mathcal{S}(p, r) \mathbf{P}_1^T$ for some $\boldsymbol{\theta} \in \mathcal{D}(p, r)$ since $\Sigma(\boldsymbol{\theta}) \in \mathcal{S}(p, r)$. Hence, we conclude that $\mathbf{x}_{\mathbf{P}_1}(\mathcal{U}) = \mathbf{P}_1 \mathcal{S}(p, r) \mathbf{P}_1^T$.

Similarly, if $\Pi' \in \{\mathbf{P}_2 \mathbf{U} \mathbf{U}^T \mathbf{P}_2^T : \mathbf{U} \in \mathbb{O}_+(p, r)\}$, then there exists some $\mathbf{U}' \in \mathbb{O}_+(p, r)$ such that $\Pi' = \mathbf{P}_2 \mathbf{U}' (\mathbf{U}')^T \mathbf{P}_2^T$, and there exists some $\boldsymbol{\varphi} \in \mathcal{A}$ such that $\mathbf{U}(\boldsymbol{\varphi}) = \mathbf{U}'$, implying that

$$\mathbf{y}_{\mathbf{P}_2}(\boldsymbol{\varphi}) = \mathbf{P}_2 \mathbf{U}(\boldsymbol{\varphi}) \mathbf{U}(\boldsymbol{\varphi})^T \mathbf{P}_2^T = \mathbf{P}_2 \mathbf{U}' (\mathbf{U}')^T \mathbf{P}_2^T = \Pi' \in \mathbf{y}_{\mathbf{P}_2}(\mathcal{A}).$$

Conversely, if $\Pi' \in \mathbf{y}_{\mathbf{P}_2}(\mathcal{A})$, then there exists $\boldsymbol{\varphi} \in \mathcal{A}$ such that $\Pi' = \mathbf{P}_2 \mathbf{U}(\boldsymbol{\varphi}) \mathbf{U}(\boldsymbol{\varphi})^T \mathbf{P}_2^T$. Since $\mathbf{U}(\boldsymbol{\varphi}) \in \mathbb{O}_+(p, r)$ automatically by the definition of $\mathbf{U}(\cdot)$, we conclude that

$$\mathbf{y}_{\mathbf{P}_2}(\mathcal{A}) = \{\mathbf{P}_2 \mathbf{U} \mathbf{U}^T \mathbf{P}_2^T : \mathbf{U} \in \mathbb{O}_+(p, r)\}.$$

- (iii) We show that $\mathbf{x}_{\mathbf{P}_1}(\mathcal{U})$ and $\mathbf{y}_{\mathbf{P}_2}(\mathcal{A})$ are open with regard to the Euclidean topology restricted to $\mathcal{M}(p, r)$ and $\mathcal{G}(p, r)$, respectively. Let $\mathbf{P}_1 \Sigma(\boldsymbol{\theta}) \mathbf{P}_1^T \in \mathbf{x}_{\mathbf{P}_1}(\mathcal{U})$ for some $\boldsymbol{\theta} \in \mathcal{D}(p, r)$ and $\mathbf{P}_2 \mathbf{U}(\boldsymbol{\varphi}_2) \mathbf{U}(\boldsymbol{\varphi}_2)^T \mathbf{P}_2 \in \mathbf{y}_{\mathbf{P}_2}(\mathcal{A})$ for some $\boldsymbol{\varphi}_2 \in \mathcal{A}$ with $\boldsymbol{\varphi}_2 = \text{vec}(\mathbf{A}_2)$ and $\mathbf{A}_2 \in \mathbb{R}^{(p-r) \times r}$, $\|\mathbf{A}_2\|_2 < 1$. Suppose

$\boldsymbol{\theta} = [\boldsymbol{\varphi}_1^T, \boldsymbol{\mu}^T]^T$, where $\boldsymbol{\varphi}_1 = \text{vec}(\mathbf{A}_1)$ for some $\|\mathbf{A}\|_2 < 1$, $\mathbf{A} \in \mathbb{R}^{(p-r) \times r}$. Let

$$\delta_1 = \min \left\{ \|\boldsymbol{\Sigma}(\boldsymbol{\theta})\|_2, \frac{\min_{k \in [r]} |\lambda_k^2(\boldsymbol{\Sigma}(\boldsymbol{\theta}))| (1 - \|\mathbf{A}_1\|_2^2)}{12\sqrt{2}\|\boldsymbol{\Sigma}(\boldsymbol{\theta})\|_2(1 + \|\mathbf{A}_1\|_2^2)} \right\} > 0,$$

$$\delta_2 = \frac{(1 - \|\mathbf{A}_1\|_2^2)}{8\sqrt{2}(1 + \|\mathbf{A}_1\|_2^2)} > 0,$$

and consider any $\boldsymbol{\Sigma}' \in \mathbf{P}_{\mathcal{S}}(p, r)\mathbf{P}^T$ with $\|\boldsymbol{\Sigma}' - \mathbf{P}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{P}^T\|_F < \delta_1$ and $\mathbf{Q}\mathbf{Q}^T \in \mathcal{G}(p, r)$ with $\|\mathbf{Q}\mathbf{Q}^T - \mathbf{P}_2\mathbf{U}(\boldsymbol{\varphi}_2)\mathbf{U}(\boldsymbol{\varphi}_2)^T\mathbf{P}_2^T\|_F < \delta_2$, where $\mathbf{Q} \in \mathbb{O}(p, r)$. It is sufficient to show that $\mathbf{P}_1^T\boldsymbol{\Sigma}'\mathbf{P}_1 \in \mathcal{S}(p, r)$ and $\mathbf{Q}\mathbf{Q}^T \in \mathbf{y}_{\mathbf{P}_2}(\mathcal{A})$.

For the first part, let $\mathbf{U}'\mathbf{S}'(\mathbf{U}')^T$ be the eigendecomposition of $\mathbf{P}_1^T\boldsymbol{\Sigma}'\mathbf{P}_1$, where $\mathbf{U}' \in \mathbb{O}(p, r)$ and \mathbf{S}' is the diagonal matrix of the nonzero eigenvalues of $\mathbf{P}_1^T\boldsymbol{\Sigma}'\mathbf{P}_1$. By Theorem 3 in [62], there exists $\mathbf{O} \in \mathbb{O}(r)$ such that

$$\|\mathbf{U}'\mathbf{O} - \mathbf{U}(\boldsymbol{\varphi})\|_F \leq \frac{6\sqrt{2}\|\boldsymbol{\Sigma}(\boldsymbol{\theta})\|_2\delta}{\min_{k \in [r]} |\lambda_k^2(\boldsymbol{\Sigma}(\boldsymbol{\theta}))|}.$$

Let \mathbf{Q}_1 be the top square block of \mathbf{U}' , i.e., $\mathbf{Q}_1 = \mathbf{I}_{p \times r}^T \mathbf{U}'$. It follows from Weyl's inequality for singular values that

$$|\sigma_r(\mathbf{Q}_1) - \sigma_r(\mathbf{I}_{p \times r}^T \mathbf{U}(\boldsymbol{\varphi}))| \leq \|\mathbf{Q}_1\mathbf{O} - \mathbf{I}_{p \times r}^T \mathbf{U}(\boldsymbol{\varphi})\|_2 \leq \|\mathbf{U}'\mathbf{O} - \mathbf{U}(\boldsymbol{\varphi})\|_F$$

$$\leq \frac{6\sqrt{2}\|\boldsymbol{\Sigma}(\boldsymbol{\theta})\|_2\delta}{\min_{k \in [r]} |\lambda_k^2(\boldsymbol{\Sigma}(\boldsymbol{\theta}))|} \leq \frac{1 - \|\mathbf{A}\|_2^2}{2(1 + \|\mathbf{A}\|_2^2)}.$$

Since $\sigma_r(\mathbf{I}_{p \times r}^T \mathbf{U}(\boldsymbol{\varphi})) = \sigma_r((\mathbf{I}_r - \mathbf{A}^T \mathbf{A})(\mathbf{I}_r + \mathbf{A}^T \mathbf{A})^{-1}) = (1 - \|\mathbf{A}\|_2^2)/(1 + \|\mathbf{A}\|_2^2)$, we obtain that

$$\sigma_r(\mathbf{Q}_1) \geq \frac{1 - \|\mathbf{A}\|_2^2}{2(1 + \|\mathbf{A}\|_2^2)} > 0.$$

This implies that the first r rows of \mathbf{U}' are linearly independent, and hence, there exists an orthogonal matrix $\mathbf{V}' \in \mathbb{O}(r)$ such that $\mathbf{U}'\mathbf{V}' \in \mathbb{O}_+(p, r)$, and hence,

$$\mathbf{P}_1^T\boldsymbol{\Sigma}'\mathbf{P}_1 = \mathbf{U}'\mathbf{S}'(\mathbf{U}')^T = (\mathbf{U}'\mathbf{V}')\{(\mathbf{V}')^T\mathbf{S}'\mathbf{V}'\}(\mathbf{U}'\mathbf{V}')^T \in \mathcal{S}(p, r).$$

This completes the proof that $\mathbf{P}_1\mathcal{S}(p, r)\mathbf{P}_1^T$ is open with regard to the Euclidean (Frobenius) topology restricted to $\mathcal{M}(p, r)$.

To show that $\mathbf{Q}\mathbf{Q}^T \in \mathbf{y}_{\mathbf{P}_2}(\mathcal{A})$, we apply Davis-Kahan theorem (see Theorem 2 in [62]) and obtain

$$\|\mathbf{P}_2^T\mathbf{Q}\mathbf{Q}' - \mathbf{U}(\boldsymbol{\varphi})\|_F \leq 2\sqrt{2}\|\mathbf{Q}\mathbf{Q}^T - \mathbf{P}_2\mathbf{U}(\boldsymbol{\varphi}_2)\mathbf{U}(\boldsymbol{\varphi}_2)^T\mathbf{P}_2^T\|_F$$

$$\leq 2\sqrt{2}\delta_2 < \frac{1 - \|\mathbf{A}_2\|_2^2}{4(1 + \|\mathbf{A}_2\|_2^2)} = \frac{1}{4}\sigma_r(\mathbf{I}_{p \times r}^T \mathbf{U}(\boldsymbol{\varphi}))$$

for some $\mathbf{O}' \in \mathbb{O}(r)$. By a similar argument, we have

$$\sigma_r(\mathbf{I}_{p \times r}^T \mathbf{P}_2^T \mathbf{Q}) > \frac{1}{2}\sigma_r(\mathbf{I}_{p \times r}^T \mathbf{Q}) > 0,$$

and hence, there exists some $\mathbf{V}'' \in \mathbb{O}(r)$ such that $\mathbf{P}_2^T \mathbf{Q}\mathbf{V}'' \in \mathbb{O}_+(p, r)$. Therefore,

$$\mathbf{Q}\mathbf{Q}^T = \mathbf{P}_2(\mathbf{P}_2^T \mathbf{Q}\mathbf{V}'')(\mathbf{P}_2^T \mathbf{Q}\mathbf{V}'')^T \mathbf{P}_2^T \in \mathbf{y}_{\mathbf{P}_2}(\mathcal{A}).$$

This shows that $\mathbf{y}_{\mathbf{P}_2}(\mathcal{A})$ is open with respect to the Euclidean topology restricted to $\mathcal{G}(p, r)$.

(iv) We show that $\mathbf{x}_{\mathbf{P}_1} : \mathcal{U} \rightarrow \mathbf{x}_{\mathbf{P}_1}(\mathcal{U})$ and $\mathbf{y}_{\mathbf{P}_2} : \mathcal{A} \rightarrow \mathbf{y}_{\mathbf{P}_2}(\mathcal{A})$ are homeomorphisms. This is equivalent to show that $\mathbf{x}_{\mathbf{P}_1}$ and $\mathbf{y}_{\mathbf{P}_2}$ are continuous and they have continuous inverses $\mathbf{x}_{\mathbf{P}_1}^{-1} : \mathbf{x}_{\mathbf{P}_1}(\mathcal{U}) \rightarrow \mathcal{U}$ and $\mathbf{y}_{\mathbf{P}_2}^{-1} : \mathbf{y}_{\mathbf{P}_2}(\mathcal{A}) \rightarrow \mathcal{A}$, respectively. Now suppose $\boldsymbol{\theta} = [\theta_1, \dots, \theta_d]^\top$ and $\boldsymbol{\varphi} = [\phi_1, \dots, \phi_s]^\top$, $s = (p-r)r$, and write

$$\begin{aligned} \text{vec}\{\mathbf{x}_{\mathbf{P}_1}(\boldsymbol{\theta})\} &= [\sigma_1(\theta_1, \dots, \theta_d), \dots, \sigma_{p^2}(\theta_1, \dots, \theta_d)]^\top, \\ \text{vec}\{\mathbf{y}_{\mathbf{P}_2}(\boldsymbol{\varphi})\} &= [\pi_1(\phi_1, \dots, \phi_s), \dots, \pi_{p^2}(\phi_1, \dots, \phi_s)]^\top. \end{aligned}$$

By Theorem 3.6 and the fact that

$$\begin{aligned} \text{vec}\{\mathbf{x}_{\mathbf{P}}(\boldsymbol{\theta})\} &= (\mathbf{P} \otimes \mathbf{P})\text{vec}\{\boldsymbol{\Sigma}(\boldsymbol{\theta})\}, \\ \text{vec}\{\mathbf{y}_{\mathbf{P}}(\boldsymbol{\varphi})\} &= (\mathbf{P} \otimes \mathbf{P})\text{vec}\{\mathbf{U}(\boldsymbol{\varphi})\mathbf{U}(\boldsymbol{\varphi})^\top\}, \end{aligned}$$

we may assume the Jacobian matrices

$$\begin{bmatrix} \frac{\partial \sigma_1}{\partial \theta_1} & \cdots & \frac{\partial \sigma_1}{\partial \theta_d} \\ \vdots & & \vdots \\ \frac{\partial \sigma_d}{\partial \theta_1} & \cdots & \frac{\partial \sigma_d}{\partial \theta_d} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \frac{\partial \pi_1}{\partial \phi_1} & \cdots & \frac{\partial \pi_1}{\partial \phi_s} \\ \vdots & & \vdots \\ \frac{\partial \pi_s}{\partial \phi_1} & \cdots & \frac{\partial \pi_s}{\partial \phi_s} \end{bmatrix}$$

are non-singular without loss of generality. Consider extensions of $\text{vec}\{\mathbf{x}_{\mathbf{P}_1}(\cdot)\}$ and $\text{vec}\{\mathbf{y}_{\mathbf{P}_2}(\cdot)\}$ defined by

$$\begin{aligned} \mathbf{g} : \mathcal{D}(p, r) \times \mathbb{R}^{p^2-d} &\rightarrow \mathbb{R}^{p^2}, \\ \mathbf{h} : \mathcal{A} \times \mathbb{R}^{p^2-s} &\rightarrow \mathbb{R}^{p^2}, \end{aligned}$$

where

$$\mathbf{g}(\theta_1, \dots, \theta_d, \theta_{d+1}, \dots, \theta_{p^2}) = \begin{bmatrix} \sigma_1(\theta_1, \dots, \theta_d) \\ \vdots \\ \sigma_d(\theta_1, \dots, \theta_d) \\ \sigma_{d+1}(\theta_1, \dots, \theta_d) + \theta_{d+1} \\ \vdots \\ \sigma_{p^2}(\theta_1, \dots, \theta_d) + \theta_{p^2} \end{bmatrix}$$

and

$$\mathbf{h}(\phi_1, \dots, \phi_s, \phi_{s+1}, \dots, \phi_{p^2}) = \begin{bmatrix} \pi_1(\phi_1, \dots, \phi_s) \\ \vdots \\ \pi_s(\phi_1, \dots, \phi_s) \\ \pi_{s+1}(\phi_1, \dots, \phi_s) + \phi_{s+1} \\ \vdots \\ \pi_{p^2}(\phi_1, \dots, \phi_s) + \phi_{p^2} \end{bmatrix}$$

Then the Jacobian matrix of \mathbf{g} with respect to the vector

$$[\theta_1, \dots, \theta_d, \theta_{d+1}, \dots, \theta_{p^2}]^\top$$

is also non-singular at $[\boldsymbol{\theta}^T, \theta_{d+1}, \dots, \theta_{p^2}]^T$ for any $[\theta_{d+1}, \dots, \theta_{p^2}]^T \in \mathbb{R}^{p^2-d}$. Similarly, the Jacobian matrix of \mathbf{h} with respect to the vector

$$[\phi_1, \dots, \phi_s, \phi_{s+1}, \dots, \phi_{p^2}]^T$$

is also non-singular at $[\boldsymbol{\varphi}^T, \phi_{s+1}, \dots, \phi_{p^2}]^T$ for any $[\phi_{s+1}, \dots, \phi_{p^2}]^T \in \mathbb{R}^{p^2-s}$. By the inverse mapping theorem, the function

$$\mathbf{g} : \mathcal{D}(p, r) \times \mathbb{R}^{p^2-d} \rightarrow \mathbb{R}^{p^2},$$

$$[\theta_1, \dots, \theta_d, \theta_{d+1}, \dots, \theta_{p^2}]^T \mapsto \mathbf{g}(\theta_1, \dots, \theta_d, \theta_{d+1}, \dots, \theta_{p^2})$$

has a continuous inverse

$$\mathbf{g}^{-1} : [\tau_1, \dots, \tau_d, \tau_{d+1}, \dots, \tau_{p^2}]^T \mapsto [\theta_1, \dots, \theta_d, \theta_{d+1}, \dots, \theta_{p^2}]^T,$$

where $\tau_k = \sigma_k(\theta_1, \dots, \theta_d)$ for all $k \in [d]$ and $\tau_k = \sigma_k(\theta_1, \dots, \theta_d) + \theta_k$ for all $k \in \{d+1, \dots, p^2\}$. Similarly, the function

$$\mathbf{h} : \mathcal{A} \times \mathbb{R}^{p^2-s} \rightarrow \mathbb{R}^{p^2},$$

$$[\phi_1, \dots, \phi_s, \phi_{s+1}, \dots, \phi_{p^2}]^T \mapsto \mathbf{h}(\phi_1, \dots, \phi_s, \phi_{s+1}, \dots, \phi_{p^2})$$

has a continuous inverse

$$\mathbf{h}^{-1} : [\alpha_1, \dots, \alpha_s, \alpha_{s+1}, \dots, \alpha_{p^2}]^T \mapsto [\phi_1, \dots, \phi_s, \phi_{s+1}, \dots, \phi_{p^2}]^T,$$

where $\alpha_k = \pi_k(\phi_1, \dots, \phi_s)$ for all $k \in [s]$ and $\alpha_k = \pi_k(\phi_1, \dots, \phi_s) + \phi_k$ for all $k \in \{s+1, \dots, p^2\}$. For any integer $m \in [p^2]$, let $\boldsymbol{\pi}_m : \mathbb{R}^{p^2} \rightarrow \mathbb{R}^d$ be the projection onto the first m coordinates, i.e.,

$$\boldsymbol{\pi}(\theta_1, \dots, \theta_m, \theta_{m+1}, \dots, \theta_{p^2}) = [\theta_1, \dots, \theta_m]^T.$$

Then we claim that $\boldsymbol{\pi}_d \circ \mathbf{g}^{-1}|_{\text{vec}(\mathbf{x}_{\mathbf{P}_1})(\mathcal{U})}$ is the inverse of $\text{vec}\{\mathbf{x}_{\mathbf{P}_1}(\cdot)\}$, and $\boldsymbol{\pi}_s \circ \mathbf{h}^{-1}|_{\text{vec}(\mathbf{y}_{\mathbf{P}_2})(\mathcal{A})}$ is the inverse of $\text{vec}\{\mathbf{y}_{\mathbf{P}_2}(\cdot)\}$. In fact, for any $\boldsymbol{\theta} = [\theta_1, \dots, \theta_d]^T \in \mathcal{D}(p, r)$ and $\boldsymbol{\varphi} = [\phi_1, \dots, \phi_s]^T \in \mathcal{A}$, by the facts that

$$\text{vec}\{\mathbf{x}_{\mathbf{P}_1}(\boldsymbol{\theta})\} = \mathbf{g}(\theta_1, \dots, \theta_d, 0, \dots, 0),$$

$$\text{vec}\{\mathbf{y}_{\mathbf{P}_2}(\boldsymbol{\varphi})\} = \mathbf{h}(\phi_1, \dots, \phi_s, 0, \dots, 0),$$

we have

$$\begin{aligned} (\boldsymbol{\pi}_d \circ \mathbf{g}^{-1}|_{\text{vec}(\mathbf{x}_{\mathbf{P}_1})(\mathcal{U})}) \circ \mathbf{x}_{\mathbf{P}_1}(\boldsymbol{\theta}) &= \{\boldsymbol{\pi}_d \circ \mathbf{g}^{-1} \circ \text{vec}(\mathbf{x}_{\mathbf{P}_1})\}(\boldsymbol{\theta}) \\ &= (\boldsymbol{\pi}_d \circ \mathbf{g}^{-1} \circ \mathbf{g})(\theta_1, \dots, \theta_d, 0, \dots, 0) \\ &= \boldsymbol{\pi}_d(\theta_1, \dots, \theta_d, 0, \dots, 0) = \boldsymbol{\theta}, \\ (\boldsymbol{\pi}_s \circ \mathbf{h}^{-1}|_{\text{vec}(\mathbf{y}_{\mathbf{P}_2})(\mathcal{A})}) \circ \mathbf{y}_{\mathbf{P}_2}(\boldsymbol{\varphi}) &= \{\boldsymbol{\pi}_s \circ \mathbf{h}^{-1} \circ \text{vec}(\mathbf{y}_{\mathbf{P}_2})\}(\boldsymbol{\varphi}) \\ &= (\boldsymbol{\pi}_s \circ \mathbf{h}^{-1} \circ \mathbf{h})(\phi_1, \dots, \phi_s, 0, \dots, 0) \\ &= \boldsymbol{\pi}_s(\phi_1, \dots, \phi_s, 0, \dots, 0) = \boldsymbol{\varphi}. \end{aligned}$$

Also, because $\boldsymbol{\pi}_d$, $\boldsymbol{\pi}_s$, \mathbf{g}^{-1} , and \mathbf{h}^{-1} are continuous, it follows that the composite restrictions $\boldsymbol{\pi}_d \circ \mathbf{g}^{-1}|_{\text{vec}(\mathbf{x}_{\mathbf{P}_1})(\mathcal{U})}$ and $\boldsymbol{\pi}_s \circ \mathbf{h}^{-1}|_{\text{vec}(\mathbf{y}_{\mathbf{P}_2})(\mathcal{A})}$ are also continuous. This shows that $\boldsymbol{\pi}_d \circ \mathbf{g}^{-1}|_{\text{vec}(\mathbf{x}_{\mathbf{P}_1})(\mathcal{U})}$ and $\boldsymbol{\pi}_s \circ \mathbf{h}^{-1}|_{\text{vec}(\mathbf{y}_{\mathbf{P}_2})(\mathcal{A})}$ are the continuous inverses of $\text{vec}\{\mathbf{x}_{\mathbf{P}_1}(\cdot)\}$ and $\text{vec}\{\mathbf{y}_{\mathbf{P}_2}(\cdot)\}$. It is also clear that $\mathbf{x}_{\mathbf{P}_1}$ and $\mathbf{y}_{\mathbf{P}_2}$ are continuous. Hence, we conclude that $\mathbf{x}_{\mathbf{P}_1} : \mathcal{U} \rightarrow \mathbf{x}_{\mathbf{P}_1}(\mathcal{U})$ and $\mathbf{y}_{\mathbf{P}_2} : \mathcal{A} \rightarrow \mathbf{y}_{\mathbf{P}_2}(\mathcal{A})$ are homeomorphisms.

(v) We show that Jacobians $\partial \text{vec}\{\mathbf{x}_{\mathbf{P}_1}(\boldsymbol{\theta})\}/\partial \boldsymbol{\theta}^T$ and $\partial \text{vec}\{\mathbf{y}_{\mathbf{P}_2}(\boldsymbol{\varphi})\}/\partial \boldsymbol{\varphi}^T$ have ranks d and s , respectively. This is immediate by Theorem 3.6 and the fact that

$$\begin{aligned}\text{vec}\{\mathbf{x}_{\mathbf{P}_1}(\boldsymbol{\theta})\} &= (\mathbf{P}_1 \otimes \mathbf{P}_1) \text{vec}\{\boldsymbol{\Sigma}(\boldsymbol{\theta})\}, \\ \text{vec}\{\mathbf{y}_{\mathbf{P}_2}(\boldsymbol{\varphi})\} &= (\mathbf{P}_2 \otimes \mathbf{P}_2) \text{vec}\{\mathbf{U}(\boldsymbol{\varphi})\mathbf{U}(\boldsymbol{\varphi})^T\}.\end{aligned}$$

Combining the results above completes the joint proof of Propositions 3.9 and 4.1. \square

Derivation of Example 4.3. To verify that $\gamma(\cdot)$ given above is indeed a geodesic, we consider a local coordinate system $(\mathcal{A}, \mathbf{y}_{\mathbf{P}})$ with $\mathbf{y}_{\mathbf{P}}(\boldsymbol{\varphi}) = \mathbf{u}\mathbf{u}^T$, $\boldsymbol{\varphi} \in \mathbb{R}^{p-1}$, $\|\boldsymbol{\varphi}\|_2 < 1$, and \mathbf{P} is a permutation matrix such that $\mathbf{P}^T \mathbf{u} \mathbf{v} \in \mathbb{O}_+(p, 1)$, where $v \in \{\pm 1\}$ adjusts the sign of the first coordinate of $\mathbf{P}^T \mathbf{u}$. Without loss of generality, we may assume that $v = +1$ because v does not affect the formula of $\gamma(\cdot)$. Since $\boldsymbol{\Xi} \in T_{\Pi} \mathcal{G}(p, 1)$, it can be written as $\boldsymbol{\Xi} = \mathbf{U}_{\perp} \mathbf{z} \mathbf{u}^T + \mathbf{u} \mathbf{z}^T \mathbf{U}_{\perp}^T$ for some $\mathbf{z} \in \mathbb{R}^{p-1}$, where $\mathbf{U}_{\perp} \in \mathbb{O}(p, p-1)$ spans the orthogonal complement of $\text{Span}(\mathbf{u})$. Let $\sigma := \|\boldsymbol{\Xi} \mathbf{u}\|_2$, $\dot{\mathbf{u}} = \boldsymbol{\Xi} \mathbf{u} / \sigma$ and write $\mathbf{P}^T \mathbf{u}$, $\mathbf{P}^T \dot{\mathbf{u}}$ in the following block form

$$\mathbf{P}^T \mathbf{u} = \begin{bmatrix} u_1 \\ \mathbf{u}_2 \end{bmatrix}, \quad \mathbf{P}^T \dot{\mathbf{u}} = \begin{bmatrix} \delta_1 \\ \boldsymbol{\delta}_2 \end{bmatrix},$$

where $u_1, \delta_1 \in \mathbb{R}$, $u_1 > 0$, and $\mathbf{u}_2, \boldsymbol{\delta}_2 \in \mathbb{R}^{p-1}$. Consider a curve $\boldsymbol{\varphi}(\cdot) : (-\epsilon, \epsilon) \rightarrow \mathcal{A} = \{\boldsymbol{\varphi} \in \mathbb{R}^{p-1} : \|\boldsymbol{\varphi}\|_2 < 1\}$ given by

$$\boldsymbol{\varphi}(t) = \frac{\mathbf{u}_2 \cos \sigma t + \boldsymbol{\delta}_2 \sin \sigma t}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t}.$$

Using the fact that $\mathbf{u}^T \mathbf{u} = \dot{\mathbf{u}}^T \dot{\mathbf{u}} = 1$ and $\mathbf{u}^T \dot{\mathbf{u}} = 0$, we obtain from a simple algebra that

$$\frac{1 - \boldsymbol{\varphi}(t)^T \boldsymbol{\varphi}(t)}{1 + \boldsymbol{\varphi}(t)^T \boldsymbol{\varphi}(t)} = u_1 \cos \sigma t + \delta_1 \sin \sigma t, \quad \frac{2\boldsymbol{\varphi}(t)}{1 + \boldsymbol{\varphi}(t)^T \boldsymbol{\varphi}(t)} = \mathbf{u}_2 \cos \sigma t + \boldsymbol{\delta}_2 \sin \sigma t.$$

It follows that

$$\begin{aligned}\mathbf{y}_{\mathbf{P}}(\boldsymbol{\varphi}(t)) &= \mathbf{P} \begin{bmatrix} u_1 \cos \sigma t + \delta_1 \sin \sigma t \\ \mathbf{u}_2 \cos \sigma t + \boldsymbol{\delta}_2 \sin \sigma t \end{bmatrix} \begin{bmatrix} u_1 \cos \sigma t + \delta_1 \sin \sigma t & \mathbf{u}_2^T \cos \sigma t + \boldsymbol{\delta}_2^T \sin \sigma t \end{bmatrix} \mathbf{P}^T \\ &= (\mathbf{u} \cos \sigma t + \dot{\mathbf{u}} \sin \sigma t)(\mathbf{u} \cos \sigma t + \dot{\mathbf{u}} \sin \sigma t)^T = \gamma(t).\end{aligned}$$

It is now sufficient to verify that $\boldsymbol{\varphi}(t)$ satisfies the geodesic equation under the local coordinate $(\mathcal{A}, \mathbf{y}_{\mathbf{P}})$. By Propositions 4.2, the Gram matrix of the Riemannian metric at $\Pi = \mathbf{y}_{\mathbf{P}}(\boldsymbol{\varphi})$ under $(\mathcal{A}, \mathbf{y}_{\mathbf{P}})$ is

$$\begin{aligned}[\mathbf{g}_{ij}]_{(p-1) \times (p-1)} &= \left\{ \frac{1}{1 + \boldsymbol{\varphi}^T \boldsymbol{\varphi}} \left(\mathbf{I}_{p-1} - \frac{\boldsymbol{\varphi} \boldsymbol{\varphi}^T}{1 + \boldsymbol{\varphi}^T \boldsymbol{\varphi}} \right) + \frac{\boldsymbol{\varphi} \boldsymbol{\varphi}^T}{(1 + \boldsymbol{\varphi}^T \boldsymbol{\varphi})^2} \right\}^2 \\ &= \frac{\mathbf{I}_{p-1}}{(1 + \boldsymbol{\varphi}^T \boldsymbol{\varphi})^2}.\end{aligned}$$

Let $[\mathbf{g}^{ij}]_{(p-1) \times (p-1)}$ be the inverse of $[\mathbf{g}_{ij}]_{(p-1) \times (p-1)}$ and write $\boldsymbol{\varphi} = [\phi_1, \dots, \phi_{p-1}]^T$. The Christoffel symbol Γ_{ij}^m for any $i, j, m \in [p-1]$ can be computed as

$$\begin{aligned}\Gamma_{ij}^m &= \frac{1}{2} \sum_{k=1}^{p-1} \left(\frac{\partial g_{jk}}{\partial \phi_i} + \frac{\partial g_{ik}}{\partial \phi_j} - \frac{\partial g_{ij}}{\partial \phi_k} \right) g^{km} \\ &= \mathbb{1}(i = j) \frac{2\phi_m}{1 + \boldsymbol{\varphi}^T \boldsymbol{\varphi}} - \mathbb{1}(i = m) \frac{2\phi_j}{1 + \boldsymbol{\varphi}^T \boldsymbol{\varphi}} - \mathbb{1}(j = m) \frac{2\phi_i}{1 + \boldsymbol{\varphi}^T \boldsymbol{\varphi}}.\end{aligned}$$

1427 Write $\boldsymbol{\varphi}(t) = [\phi_1(t), \dots, \phi_{p-1}(t)]^T$. We now argue that for each $m \in [p-1]$, $\phi_m(t)$
 1428 satisfies the geodesic equation

$$1429 \quad \phi_m''(t) + \sum_{i=1}^{p-1} \sum_{j=1}^{p-1} \Gamma_{ij}^n \phi_i'(t) \phi_j'(t) = 0.$$

1430 Write $\mathbf{u}_2 = [u_{21}, \dots, u_{2(p-1)}]^T$ and $\boldsymbol{\delta}_2 = [\delta_{21}, \dots, \delta_{2(p-1)}]^T$. A direct computation
 1431 yields

$$1432 \quad \phi_m'(t) = \frac{-\sigma u_{2m} \sin \sigma t + \sigma \delta_{2m} \cos \sigma t}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} - \frac{(u_{2m} \cos \sigma t + \delta_{2m} \sin \sigma t)(-\sigma u_1 \sin \sigma t + \sigma \delta_1)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2},$$

1434 and

$$1435 \quad \frac{2\phi_i(t)\phi_i'(t)}{1 + \boldsymbol{\varphi}(t)^T \boldsymbol{\varphi}(t)} = \frac{(-\sigma u_{2i}^2 + \sigma \delta_{2i}^2) \sin \sigma t \cos \sigma t + \sigma u_{2i} \delta_{2i} (\cos^2 \sigma t - \sin^2 \sigma t)}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t}$$

$$1436 \quad - \frac{(u_{2i}^2 \cos^2 \sigma t + \delta_{2i}^2 \sin^2 \sigma t + 2u_{2i} \delta_{2i} \sin \sigma t \cos \sigma t)(-\sigma u_1 \sin \sigma t + \sigma \delta_1)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2}.$$

1438 It follows that

$$1439 \quad \sum_{i=1}^{p-1} \frac{2\phi_i(t)\phi_i'(t)}{1 + \boldsymbol{\varphi}(t)^T \boldsymbol{\varphi}(t)}$$

$$1440 \quad = \frac{(-\sigma + \sigma u_1^2 + \sigma - \sigma \delta_1^2) \sin \sigma t \cos \sigma t - \sigma u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t)}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t}$$

$$1441 \quad - \frac{(1 - u_1^2 \cos^2 \sigma t - \delta_1^2 \sin^2 \sigma t - 2u_1 \delta_1 \sin \sigma t \cos \sigma t)(-\sigma u_1 \sin \sigma t + \sigma \delta_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2}$$

$$1442 \quad = \frac{(\sigma u_1^2 - \sigma \delta_1^2) \sin \sigma t \cos \sigma t - \sigma u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t)}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t}$$

$$1443 \quad - \frac{(1 - u_1 \cos \sigma t - \delta_1^2 \sin^2 \sigma t)(-\sigma u_1 \sin \sigma t + \sigma \delta_1 \cos \sigma t)}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t}$$

$$1444 \quad = \frac{(\sigma u_1^2 - \sigma \delta_1^2) \sin \sigma t \cos \sigma t - \sigma u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t) + \sigma u_1 \sin \sigma t - \sigma \delta_1 \cos \sigma t}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t}$$

$$1445 \quad + \frac{-\sigma u_1^2 \cos \sigma t \sin \sigma t + \sigma u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t) + \sigma u_1^2 \sin \sigma t \cos \sigma t}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t}$$

$$1446 \quad = \frac{\sigma u_1 \sin \sigma t - \sigma \delta_1 \cos \sigma t}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t}.$$

1448 Therefore,

$$1449 \quad (A.1) \quad 2\phi_m'(t) \sum_{i=1}^{p-1} \frac{2\phi_i(t)\phi_i'(t)}{1 + \boldsymbol{\varphi}(t)^T \boldsymbol{\varphi}(t)}$$

$$= \frac{-2(\sigma u_{2m} \sin \sigma t - \sigma \delta_{2m} \cos \sigma t)(\sigma u_1 \sin \sigma t - \sigma \delta_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2}$$

$$+ \frac{2(u_{2m} \cos \sigma t + \delta_{2m} \sin \sigma t)(\sigma u_1 \sin \sigma t - \sigma \delta_1 \cos \sigma t)^2}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^3}.$$

Also, another direct computation shows that

$$\begin{aligned}
& \sum_{i=1}^{p-1} \frac{2\{\phi'_i(t)\}^2}{1 + \boldsymbol{\varphi}(t)^T \boldsymbol{\varphi}(t)} \\
&= \frac{\sigma^2(1 - u_1^2) \sin^2 \sigma t + \sigma^2(1 - \delta_1^2) \cos^2 \sigma t + 2\sigma^2 u_1 \delta_1 \sin \sigma t \cos \sigma t}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} \\
&+ \frac{\sigma^2 u_1^2 \sin^2 \sigma t + \sigma^2 \delta_1^2 \cos^2 \sigma t - 2\sigma^2 u_1 \delta_1 \sin \sigma t \cos \sigma t}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^3} \{1 - (u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2\} \\
&+ \frac{2(\sigma u_1 \sin \sigma t - \sigma \delta_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} \{\sigma(u_1^2 - \delta_1^2) \sin \sigma t \cos \sigma t - \sigma u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t)\} \\
&= \frac{\sigma^2 \{1 - (u_1 \sin \sigma t - \delta_1 \cos \sigma t)^2\}}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} \\
&+ \frac{\sigma^2 (u_1 \sin \sigma t - \delta_1 \cos \sigma t)^2 (1 - u_1 \cos \sigma t - \delta_1 \sin \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} \\
&+ \frac{2(\sigma u_1 \sin \sigma t - \sigma \delta_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} \{\sigma(u_1^2 - \delta_1^2) \sin \sigma t \cos \sigma t - \sigma u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t)\} \\
&= \frac{\sigma^2 \{1 - (u_1 \sin \sigma t - \delta_1 \cos \sigma t)^2\}}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} \\
&+ \frac{\sigma^2 (u_1 \sin \sigma t - \delta_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} \\
&\times \{u_1 \sin \sigma t - \delta_1 \cos \sigma t + (\delta_1^2 - u_1^2) \sin \sigma t \cos \sigma t + u_1 \delta_1 (\sin^2 \sigma t - \cos^2 \sigma t)\} \\
&+ \frac{\sigma^2 (u_1 \sin \sigma t - \delta_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} \\
&\times \{2(u_1^2 - \delta_1^2) \sin \sigma t \cos \sigma t - 2\sigma u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t)\} \\
&= \frac{\sigma^2 \{1 - (u_1 \sin \sigma t - \delta_1 \cos \sigma t)^2\}}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} \\
&+ \frac{\sigma^2 (u_1 \sin \sigma t - \delta_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} \\
&\times \{u_1 \sin \sigma t - \delta_1 \cos \sigma t + (u_1^2 - \delta_1^2) \sin \sigma t \cos \sigma t + u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t)\} \\
&= \frac{\sigma^2 \{1 - (u_1 \sin \sigma t - \delta_1 \cos \sigma t)^2\}}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} + \frac{\sigma^2 (u_1 \sin \sigma t - \delta_1 \cos \sigma t)^2}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} \\
&= \frac{\sigma^2}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t}.
\end{aligned}$$

Hence,

$$\phi_m(t) \sum_{i=1}^{p-1} \frac{2\{\phi'_i(t)\}^2}{1 + \boldsymbol{\varphi}(t)^T \boldsymbol{\varphi}(t)} = \frac{\sigma^2 (u_{2m} \cos \sigma t + \delta_{2m} \sin \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2}.$$

Using the formula for the Christoffel symbols, the geodesic equation becomes

$$\phi_m''(t) = 2\phi_m'(t) \sum_{i=1}^{p-1} \frac{2\phi_i(t)\phi'_i(t)}{1 + \boldsymbol{\varphi}(t)^T \boldsymbol{\varphi}(t)} - \sum_{i=1}^{p-1} \frac{2\phi_m'(t)\{\phi'_i(t)\}^2}{1 + \boldsymbol{\varphi}(t)^T \boldsymbol{\varphi}(t)}.$$

1474 A direct computation shows that

$$\begin{aligned}
 \phi_m''(t) &= \frac{-\sigma^2(u_{2m} \cos \sigma t + \delta_{2m} \sin \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} \\
 &\quad - \frac{2(-\sigma u_{2m} \sin \sigma t + \sigma \delta_{2m} \cos \sigma t)(-\sigma u_1 \sin \sigma t + \sigma \delta_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} \\
 &\quad + \frac{2(u_{2m} \cos \sigma t + \delta_{2m} \sin \sigma t)(-\sigma u_1 \sin \sigma t + \sigma \delta_1 \cos \sigma t)^2}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^3}.
 \end{aligned}
 \tag{A.3}$$

1476 Hence, the verification that $\varphi(t)$ satisfies the above equation is completed by combin-
 1477 ing the results in (A.1), (A.2), and the formula for $\phi_m''(t)$ in (A.3). \square

1478 *Proof of Proposition 4.4.* Since $\Sigma(\theta) = \mathbf{D}\Sigma(\theta_0)\mathbf{D}^T$, it follows that the leading
 1479 eigenvector \mathbf{u} of $\Sigma(\theta)$ has the form $\mathbf{u} = \mathbf{D}\mathbf{u}_0 v / \|\mathbf{D}\mathbf{u}_0\|_2$, where $v \in \{\pm 1\}$ adjusts for
 1480 the sign of the first element of $\mathbf{D}\mathbf{u}_0$ such that $\mathbf{D}\mathbf{u}_0 / \|\mathbf{D}\mathbf{u}_0\|_2 \in \mathbb{O}_+(p, 1)$. Write $\mathbf{D} - \mathbf{I}_p$
 1481 in the block form

$$\mathbf{D} - \mathbf{I}_p = \begin{bmatrix} \epsilon & \boldsymbol{\alpha}^T \\ \boldsymbol{\beta} & \boldsymbol{\Gamma} \end{bmatrix},$$

1483 where $\epsilon \in \mathbb{R}$, $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^{p-1}$, and $\boldsymbol{\Gamma} \in \mathbb{R}^{(p-1) \times (p-1)}$. Because the first element of $\mathbf{D}\mathbf{u}_0$
 1484 can be written as

$$\begin{aligned}
 [1 + \epsilon \quad \boldsymbol{\alpha}^T] \mathbf{u}_0 &= \frac{(1 + \epsilon)(1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0) + 2\boldsymbol{\alpha}^T \boldsymbol{\varphi}_0}{1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0} \geq \frac{1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0}{1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0} - |\epsilon| - \|\boldsymbol{\alpha}\|_2 \\
 &\geq \frac{1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0}{1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0} - 2\|\mathbf{D} - \mathbf{I}_p\|_F \geq \frac{1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0}{2(1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0)} > 0.
 \end{aligned}$$

1488 This implies that $v = +1$ and we can remove v from the expression of \mathbf{u} . Also, observe
 1489 that

$$\mathbf{D}\mathbf{u}_0 = \begin{bmatrix} (1 + \epsilon) \left(\frac{1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0}{1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0} \right) + \frac{2\boldsymbol{\alpha}^T \boldsymbol{\varphi}_0}{1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0} \\ \boldsymbol{\beta} \left(\frac{1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0}{1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0} \right) + \frac{2(\mathbf{I}_{p-1} + \boldsymbol{\Gamma})\boldsymbol{\varphi}_0}{1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0} \end{bmatrix}.$$

1492 By the inverse Cayley parameterization (2.4), we have

$$\begin{aligned}
 \varphi &= \frac{\boldsymbol{\beta}(1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0) + 2(\mathbf{I}_{p-1} + \boldsymbol{\Gamma})\boldsymbol{\varphi}_0}{\|\mathbf{D}\mathbf{u}_0\|_2(1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0)} \left\{ 1 + \frac{(1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0)(1 + \epsilon) + 2\boldsymbol{\alpha}^T \boldsymbol{\varphi}_0}{\|\mathbf{D}\mathbf{u}_0\|_2(1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0)} \right\}^{-1} \\
 &= \frac{\boldsymbol{\beta}(1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0) + 2(\mathbf{I}_{p-1} + \boldsymbol{\Gamma})\boldsymbol{\varphi}_0}{(1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0)\|\mathbf{D}\mathbf{u}_0\|_2 + (1 + \epsilon)(1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0) + 2\boldsymbol{\alpha}^T \boldsymbol{\varphi}_0}.
 \end{aligned}$$

1496 Note that the denominator satisfies

$$\begin{aligned}
 &(1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0)\|\mathbf{D}\mathbf{u}_0\|_2 + (1 + \epsilon)(1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0) + 2\boldsymbol{\alpha}^T \boldsymbol{\varphi}_0 \\
 &= 2 + (1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0)(\|\mathbf{D}\mathbf{u}_0\|_2 - 1) + \epsilon(1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0) + 2\boldsymbol{\alpha}^T \boldsymbol{\varphi}_0.
 \end{aligned}$$

1500 It follows that

$$(1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0)\|\mathbf{D}\mathbf{u}_0\|_2 + (1 + \epsilon)(1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0) + 2\boldsymbol{\alpha}^T \boldsymbol{\varphi}_0 \geq 2 - 4\|\mathbf{D} - \mathbf{I}_p\|_F$$

1503 and

$$\begin{aligned}
 \|\varphi - \boldsymbol{\varphi}_0\|_2 &\leq \frac{(1 + \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0)\|\mathbf{D}\mathbf{u}_0\|_2 - 1}{2 - 4\|\mathbf{D} - \mathbf{I}_p\|_F} + \frac{(1 - \boldsymbol{\varphi}_0^T \boldsymbol{\varphi}_0)(|\epsilon| + \|\boldsymbol{\beta}\|_2)}{2 - 4\|\mathbf{D} - \mathbf{I}_p\|_F} \\
 &\quad + \frac{2\|\boldsymbol{\Gamma}\boldsymbol{\varphi}_0 - \boldsymbol{\varphi}_0\boldsymbol{\alpha}^T\|_2}{2 - 4\|\mathbf{D} - \mathbf{I}_p\|_F}.
 \end{aligned}$$

1505
 1506

\square

REFERENCES

- [1] E. ABBE, *Community detection and stochastic block models: Recent developments*, J. Mach. Learn. Res., 18 (2018), pp. 1–86.
- [2] E. ABBE, J. FAN, K. WANG, AND Y. ZHONG, *Entrywise eigenvector analysis of random matrices with low expected rank*, Ann. Statist., 48 (2020), pp. 1452–1474, <https://doi.org/10.1214/19-AOS1854>, <https://doi.org/10.1214/19-AOS1854>.
- [3] P.-A. ABSIL, R. MAHONY, AND R. SEPULCHRE, *Riemannian geometry of grassmann manifolds with a view on algorithmic computation*, Acta Appl. Math., 80 (2004), pp. 199–220.
- [4] P.-A. ABSIL, R. MAHONY, AND R. SEPULCHRE, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2009, <https://doi.org/doi:10.1515/9781400830244>, <https://doi.org/doi:10.1515/9781400830244>.
- [5] J. AGTERBERG, Z. LUBBERTS, AND C. E. PRIEBE, *Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence*, IEEE Trans. Inform. Theory, 68 (2022), pp. 4618–4650, <https://doi.org/10.1109/TIT.2022.3159085>.
- [6] A. A. AMINI AND M. J. WAINWRIGHT, *High-dimensional analysis of semidefinite relaxations for sparse principal components*, Ann. Statist., 37 (2009), pp. 2877–2921, <https://doi.org/10.1214/08-AOS664>, <https://doi.org/10.1214/08-AOS664>.
- [7] A. ATHREYA, D. E. FISHKIND, M. TANG, C. E. PRIEBE, Y. PARK, J. T. VOGELSTEIN, K. LEVIN, V. LYZINSKI, AND Y. QIN, *Statistical inference on random dot product graphs: a survey*, J. Mach. Learn. Res., 18 (2017), pp. 8393–8484.
- [8] T. BENDOKAT, R. ZIMMERMANN, AND P.-A. ABSIL, *A Grassmann manifold handbook: Basic geometry and computational aspects*, arXiv preprint arXiv:2011.13699, (2020).
- [9] R. BHATIA, *Matrix analysis*, vol. 169, Springer Science & Business Media, 1997.
- [10] J.-C. BOURIN AND E.-Y. LEE, *Unitary orbits of Hermitian operators with convex or concave functions*, Bulletin of the London Mathematical Society, 44 (2012), pp. 1085–1102, <https://doi.org/10.1112/blms/bds080>, <https://doi.org/10.1112/blms/bds080>, <https://arxiv.org/abs/https://academic.oup.com/blms/article-pdf/44/6/1085/1072920/bds080.pdf>.
- [11] T. CAI, Z. MA, AND Y. WU, *Optimal estimation and rank detection for sparse spiked covariance matrices*, Probab. Theory Related Fields, 161 (2015), pp. 781–815.
- [12] T. T. CAI, Z. MA, AND Y. WU, *Sparse PCA: Optimal rates and adaptive estimation*, Ann. Statist., 41 (2013), pp. 3074–3110.
- [13] T. T. CAI, Z. REN, AND H. H. ZHOU, *Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation*, Electron. J. Statist., 10 (2016), pp. 1–59, <https://doi.org/10.1214/15-EJS1081>, <https://doi.org/10.1214/15-EJS1081>.
- [14] T. T. CAI AND A. ZHANG, *Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics*, Ann. Statist., 46 (2018), pp. 60–89, <https://doi.org/10.1214/17-AOS1541>, <https://doi.org/10.1214/17-AOS1541>.
- [15] E. J. CANDÈS AND B. RECHT, *Exact matrix completion via convex optimization*, Found. Comput. Math., 9 (2009), pp. 717–772.
- [16] E. J. CANDÈS AND T. TAO, *The power of convex relaxation: Near-optimal matrix completion*, IEEE Trans. Inform. Theory, 56 (2010), pp. 2053–2080.
- [17] J. CAPE, M. TANG, AND C. E. PRIEBE, *Signal-plus-noise matrix models: eigenvector deviations and fluctuations*, Biometrika, 106 (2019), pp. 243–250, <https://doi.org/10.1093/biomet/asy070>, <https://arxiv.org/abs/https://academic.oup.com/biomet/article-pdf/106/1/243/27774679/asy070.pdf>.
- [18] J. CAPE, M. TANG, AND C. E. PRIEBE, *The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics*, Ann. Statist., 47 (2019), pp. 2405–2439, <https://doi.org/10.1214/18-AOS1752>, <https://doi.org/10.1214/18-AOS1752>.
- [19] X.-W. CHANG AND R.-C. LI, *Multiplicative perturbation analysis for qr factorizations*, Numer. Algebra Control Optim., 1 (2011), pp. 301–316, <https://doi.org/10.3934/naco.2011.1.301>, <https://doi.org/10.3934/naco.2011.1.301>, <https://arxiv.org/abs/https://doi.org/10.3934/naco.2011.1.301>.
- [20] S. CHATTERJEE, *Matrix estimation by universal singular value thresholding*, Ann. Statist., 43 (2015), pp. 177–214, <https://doi.org/10.1214/14-AOS1272>, <https://doi.org/10.1214/14-AOS1272>.
- [21] C. DAVIS AND W. M. KAHAN, *The rotation of eigenvectors by a perturbation. iii*, SIAM J. Numer. Anal., 7 (1970), pp. 1–46, <https://doi.org/10.1137/0707001>, <https://doi.org/10.1137/0707001>, <https://arxiv.org/abs/https://doi.org/10.1137/0707001>.
- [22] M. DO CARMO, *Riemannian Geometry*, Mathematics (Boston, Mass.), Birkhäuser, 1992, <https://books.google.com/books?id=uXJQqgAACAAJ>.
- [23] D. DONOHO AND M. GAVISH, *Minimax risk of matrix denoising by singular value thresholding*, Ann. Statist., 42 (2014), pp. 2413–2440, <https://doi.org/10.1214/14-AOS1257>, <https://doi.org/10.1214/14-AOS1257>.

- doi.org/10.1214/14-AOS1257.
- [24] D. DONOHO, M. GAVISH, AND I. JOHNSTONE, *Optimal shrinkage of eigenvalues in the spiked covariance model*, Ann. Statist., 46 (2018), pp. 1742–1778, <https://doi.org/10.1214/17-AOS1601>, <https://doi.org/10.1214/17-AOS1601>.
- [25] Z. DRMAC, *On principal angles between subspaces of euclidean space*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 173–194, <https://doi.org/10.1137/S0895479897320824>, <https://doi.org/10.1137/S0895479897320824>, <https://arxiv.org/abs/https://doi.org/10.1137/S0895479897320824>.
- [26] A. EDELMAN, T. A. ARIAS, AND S. T. SMITH, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 303–353, <https://doi.org/10.1137/S0895479895290954>, <https://doi.org/10.1137/S0895479895290954>, <https://arxiv.org/abs/https://doi.org/10.1137/S0895479895290954>.
- [27] K. EICHLER, F. LI, A. LITWIN-KUMAR, Y. PARK, I. ANDRADE, C. M. SCHNEIDER-MIZELL, T. SAUMWEBER, A. HUSER, C. ESCHBACH, B. GERBER, ET AL., *The complete connectome of a learning and memory centre in an insect brain*, Nature, 548 (2017), pp. 175–182.
- [28] J. ELDRIDGE, M. BELKIN, AND Y. WANG, *Unperturbed: spectral analysis beyond davis-kahan*, in Proceedings of Algorithmic Learning Theory, F. Janoos, M. Mohri, and K. Sridharan, eds., vol. 83 of Proceedings of Machine Learning Research, PMLR, 07–09 Apr 2018, pp. 321–358, <http://proceedings.mlr.press/v83/eldridge18a.html>.
- [29] J. FAN, Y. LIAO, AND M. MINCHEVA, *Large covariance estimation by thresholding principal orthogonal complements*, J. R. Stat. Soc. Ser. B Stat. Methodol., 75 (2013), pp. 603–680.
- [30] J. FAN, W. WANG, AND Y. ZHONG, *An ℓ_∞ eigenvector perturbation bound and its application*, J. Mach. Learn. Res., 18 (2018), pp. 1–42, <http://jmlr.org/papers/v18/16-140.html>.
- [31] J. FERRER, M. GARIA, AND F. PUERTA, *Differentiable families of subspaces*, Linear Algebra Appl., 199 (1994), pp. 229–252, [https://doi.org/https://doi.org/10.1016/0024-3795\(94\)90351-4](https://doi.org/https://doi.org/10.1016/0024-3795(94)90351-4), <https://www.sciencedirect.com/science/article/pii/0024379594903514>. Special Issue Honoring Ingram Olkin.
- [32] D. GOLDBERG, D. NICHOLS, B. M. OKI, AND D. TERRY, *Using collaborative filtering to weave an information tapestry*, Commun. ACM, 35 (1992), pp. 61–70.
- [33] I. C. F. IPSEN, *Relative perturbation results for matrix eigenvalues and singular values*, Acta Numer., 7 (1998), p. 151201, <https://doi.org/10.1017/S0962492900002828>.
- [34] M. JAUCH, P. D. HOFF, AND D. B. DUNSON, *Random orthogonal matrices and the Cayley transform*, Bernoulli, 26 (2020), pp. 1560–1586, <https://doi.org/10.3150/19-BEJ1176>, <https://doi.org/10.3150/19-BEJ1176>.
- [35] I. M. JOHNSTONE, *On the distribution of the largest eigenvalue in principal components analysis*, Ann. Statist., (2001), pp. 295–327.
- [36] V. KOLTCHINSKII AND K. LOUNICI, *New asymptotic results in principal component analysis*, Sankhya Ser A, 79 (2017), pp. 254–297.
- [37] R.-C. LI, *Relative perturbation bounds for the unitary polar factor*, BIT Numer. Math., 37 (1997), pp. 67–75.
- [38] R.-C. LI, *Relative Perturbation Theory: I. Eigenvalue and Singular Value Variations*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 956–982, <https://doi.org/10.1137/S089547989629849X>, <https://doi.org/10.1137/S089547989629849X>, <https://arxiv.org/abs/https://doi.org/10.1137/S089547989629849X>.
- [39] R.-C. LI, *Relative Perturbation Theory: II. Eigenspace and Singular Subspace Variations*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 471–492, <https://doi.org/10.1137/S0895479896298506>, <https://doi.org/10.1137/S0895479896298506>, <https://arxiv.org/abs/https://doi.org/10.1137/S0895479896298506>.
- [40] R.-C. LI, *Relative perturbation bounds for positive polar factors of graded matrices*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 424–433.
- [41] Z. MA, *Sparse principal component analysis and iterative thresholding*, Ann. Statist., 41 (2013), pp. 772–801, <https://doi.org/10.1214/13-AOS1097>, <https://doi.org/10.1214/13-AOS1097>.
- [42] J. R. MAGNUS, *Linear structures*, no. 42, Charles Griffin & Company LTD, Oxford University Press, London, 1988.
- [43] J. R. MAGNUS AND H. NEUDECKER, *The commutation matrix: Some properties and applications*, Ann. Statist., 7 (1979), pp. 381–394, <https://doi.org/10.1214/aos/1176344621>, <https://doi.org/10.1214/aos/1176344621>.
- [44] C. L. M. NICKEL, *Random dot product graphs a model for social networks*, PhD thesis, Johns Hopkins University, 2008.
- [45] S. O’ROURKE, V. VU, AND K. WANG, *Random perturbation of low rank matrices: Improving classical bounds*, Linear Algebra Appl., 540 (2018), pp. 26–59.
- [46] L. QIU, Y. ZHANG, AND C.-K. LI, *Unitarily invariant metrics on the grassmann space*, SIAM

- J. Matrix Anal. Appl., 27 (2005), pp. 507–531, <https://doi.org/10.1137/040607605>, <https://doi.org/10.1137/040607605>, <https://arxiv.org/abs/https://doi.org/10.1137/040607605>.
- [47] K. ROHE, S. CHATTERJEE, AND B. YU, *Spectral clustering and the high-dimensional stochastic blockmodel*, Ann. Statist., 39 (2011), pp. 1878–1915, <https://doi.org/10.1214/11-AOS887>, <https://doi.org/10.1214/11-AOS887>.
- [48] Y. SHKOLNISKY AND A. SINGER, *Viewing direction estimation in cryo-em using synchronization*, SIAM J. Imaging Sci., 5 (2012), pp. 1088–1110, <https://doi.org/10.1137/120863642>, <https://doi.org/10.1137/120863642>, <https://arxiv.org/abs/https://doi.org/10.1137/120863642>.
- [49] G. W. STEWART AND J.-G. SUN, *Matrix Perturbation Theory*, Academic Press, 1990.
- [50] D. L. SUSSMAN, M. TANG, D. E. FISHKIND, AND C. E. PRIEBE, *A consistent adjacency spectral embedding for stochastic blockmodel graphs*, J. Amer. Statist. Assoc., 107 (2012), pp. 1119–1128.
- [51] M. TANG AND C. E. PRIEBE, *Limit theorems for eigenvectors of the normalized laplacian for random graphs*, Ann. Statist., 46 (2018), pp. 2360–2415.
- [52] R. TANG, M. KETCHA, A. BADEA, E. D. CALABRESE, D. S. MARGULIES, J. T. VOGELSTEIN, C. E. PRIEBE, AND D. L. SUSSMAN, *Connectome smoothing via low-rank approximations*, IEEE Trans. Med. Imaging, 38 (2019), pp. 1446–1456, <https://doi.org/10.1109/TMI.2018.2885968>.
- [53] V. VU, *Singular vectors under random perturbation*, Random Structures & Algorithms, 39 (2011), pp. 526–538, <https://doi.org/https://doi.org/10.1002/rsa.20367>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20367>, <https://arxiv.org/abs/https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.20367>.
- [54] V. VU AND J. LEI, *Minimax rates of estimation for sparse pca in high dimensions*, in Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, N. D. Lawrence and M. Girolami, eds., vol. 22 of Proceedings of Machine Learning Research, La Palma, Canary Islands, 21–23 Apr 2012, PMLR, pp. 1278–1286, <http://proceedings.mlr.press/v22/vu12.html>.
- [55] V. Q. VU AND J. LEI, *Minimax sparse principal subspace estimation in high dimensions*, Ann. Statist., 41 (2013), pp. 2905–2947.
- [56] P.-Å. WEDIN, *Perturbation bounds in connection with singular value decomposition*, 12 (1972), pp. 99–111.
- [57] P. Å. WEDIN, *On angles between subspaces of a finite dimensional inner product space*, in Matrix Pencils, B. Kågström and A. Ruhe, eds., Berlin, Heidelberg, 1983, Springer Berlin Heidelberg, pp. 263–285.
- [58] Y.-C. WONG, *Differential geometry of grassmann manifolds*, Proceedings of the National Academy of Sciences, 57 (1967), pp. 589–594, <https://doi.org/10.1073/pnas.57.3.589>, <https://www.pnas.org/doi/abs/10.1073/pnas.57.3.589>, <https://arxiv.org/abs/https://www.pnas.org/doi/pdf/10.1073/pnas.57.3.589>.
- [59] F. XIE, *Entrywise limit theorems of eigenvectors for signal-plus-noise matrix models with weak signals*, arXiv preprint arXiv:2106.09840, (2022).
- [60] F. XIE AND Y. XU, *Optimal Bayesian estimation for random dot product graphs*, Biometrika, 107 (2020), pp. 875–889, <https://doi.org/10.1093/biomet/asaa031>, <https://doi.org/10.1093/biomet/asaa031>, <https://arxiv.org/abs/https://academic.oup.com/biomet/article-pdf/107/4/875/34865852/asaa031.pdf>.
- [61] S. J. YOUNG AND E. R. SCHEINERMAN, *Random dot product graph models for social networks*, in International Workshop on Algorithms and Models for the Web-Graph, Springer, 2007, pp. 138–149.
- [62] Y. YU, T. WANG, AND R. J. SAMWORTH, *A useful variant of the DavisKahan theorem for statisticians*, Biometrika, 102 (2014), pp. 315–323, <https://doi.org/10.1093/biomet/asv008>, <https://doi.org/10.1093/biomet/asv008>, <https://arxiv.org/abs/https://academic.oup.com/biomet/article-pdf/102/2/315/9642505/asv008.pdf>.
- [63] F. ZHANG AND Q. ZHANG, *Eigenvalue inequalities for matrix product*, IEEE Trans. Automat. Control, 51 (2006), pp. 1506–1509, <https://doi.org/10.1109/TAC.2006.880787>.
- [64] L. ZHANG AND R. LI, *Maximization of the sum of the trace ratio on the stiefel manifold, ii: Computation*, Sci. China Math., 58 (2015), pp. 1549–1566.
- [65] H. ZOU, T. HASTIE, AND R. TIBSHIRANI, *Sparse principal component analysis*, J. Comput. Graph. Statist., 15 (2006), pp. 265–286, <https://doi.org/10.1198/106186006X113430>, <https://doi.org/10.1198/106186006X113430>, <https://arxiv.org/abs/https://doi.org/10.1198/106186006X113430>.