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

FANGZHENG XIE†

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

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

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

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

Spectral methods have been ubiquitous to gain insight into low-rank matrix models in the presence of high-dimensional data. For example, in the stochastic block model, the $K$-means clustering procedure is applied to the rows of the leading eigenvector matrix of the observed adjacency matrix or its normalized Laplacian matrix to discover the underlying community structure [1, 47, 50, 51]. Meanwhile, the theoretical understanding of spectral methods has also been developed based on matrix perturbation analysis [9, 14, 21, 49, 56, 62]. Specifically, given an approximately low-rank matrix $\Sigma_0$ and a perturbation matrix $E$ that is comparatively smaller than $\Sigma_0$ in magnitude, matrix perturbation analysis studies how the eigenspaces or singular subspaces of the perturbed matrix $\Sigma := \Sigma_0 + E$ differ from those of the original matrix $\Sigma_0$ in terms of the behavior of the perturbation $E$. Notably, when $\Sigma$ and $\Sigma_0$ are symmetric matrices, the famous Davis-Kahan theorem [21, 62] asserts that the distance between the subspace spanned by the leading eigenvectors of $\Sigma$ and that of $\Sigma_0$, formally defined via the canonical angles (see Section 2.1 below), can be upper bounded by the matrix norms of $E$. Several extensions and generalizations of the matrix perturbation tools have been developed. In [14, 56], the authors later extended 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).

This manuscript is for review purposes only.
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 \( \Sigma_0 \) 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 \( \Sigma \) is the same as the rank of \( \Sigma_0 \). This is slightly different from the classical random perturbation setup where the referential matrix \( \Sigma_0 \) is perturbed by a mean-zero but potentially full-rank random matrix \( \mathbf{E} \), as the resulting perturbed matrix \( \Sigma \) 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 $A^\dagger$ denotes the Moore-Penrose pseudoinverse of an arbitrary matrix $A$. For any two positive semidefinite matrices $A$ and $B$ of the same dimension, we use $A \succeq B$ ($A \preceq B$, resp.) to indicate that $A - B$ is positive semidefinite ($B - A$ is positive semidefinite, resp.). The $r \times r$ identity matrix is denoted by $I_r$, 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 $O(p, r) = \{ U \in \mathbb{R}^{p \times r} : U^T U = I_r \}$ the collection of all $p \times r$ Stiefel matrices and write $O(r, r) := O(r, r)$. For any $Y \in \mathbb{R}^{r \times r}$, we use $\text{Span}(Y)$ to denote the $r$-dimensional subspace in $\mathbb{R}^p$ spanned by the columns of $Y$. The collection of all $r \times r$ symmetric matrices is denoted by $M(r)$ and the collection of all $r \times r$ symmetric positive definite matrices is denoted by $M_+(r)$.

For a matrix $\Sigma \in \mathbb{R}^{p \times p}$ and indices $i \in [p_1], j \in [p_2]$, let $[\Sigma]_{ij}$ denote the element on the $i$th row and $j$th column of $\Sigma$, $[\Sigma]_{i*}$ denote the $i$th row of $\Sigma$, and $[\Sigma]_{*j}$ denote the $j$th column of $\Sigma$. Furthermore, we use $\sigma_1(\Sigma), \ldots, \sigma_{pq}(\Sigma)$ to denote the singular values of $\Sigma$ sorted in the non-increasing order, i.e., $\sigma_1(\Sigma) \geq \ldots \geq \sigma_{pq}(\Sigma)$. When $\Sigma$ is a $p \times p$ symmetric square matrix, $\lambda_1(\Sigma), \ldots, \lambda_p(\Sigma)$ denote the eigenvalues of $\Sigma$ sorted in the non-increasing order in magnitude, namely, $|\lambda_1(\Sigma)| \geq \ldots \geq |\lambda_p(\Sigma)|$.

The spectral norm of a general matrix $\Sigma$, denoted by $\|\Sigma\|_2$, is the largest singular value of $\Sigma$, and the Frobenius norm of $\Sigma$, denoted by $\|\Sigma\|_F$, is defined to be $\|\Sigma\|_F = \left(\sum_{i=1}^p \sum_{j=1}^p |\Sigma_{i j}|^2\right)^{1/2}$. For a Euclidean vector $x = [x_1, \ldots, x_p]^T \in \mathbb{R}^p$, we denote $|x|_i := x_i$, $\|x\|_2$ the usual Euclidean norm $\|x\|_2 = \left(\sum_i x_i^2\right)^{1/2}$, let $B_2(x, \epsilon) := \{ y \in \mathbb{R}^p : \|y - x\|_2 < \epsilon \}$, and let $\text{diag}(x)$ 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 $M$, the operator vec$(\cdot)$ transform $M$ to a $p_1 p_2$-dimensional Euclidean vector by stacking the columns of $\Sigma$ consecutively, i.e.,

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

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

$$\text{vech}(M) = [\{M\}_{11}, \{M\}_{21}, \ldots, \{M\}_{r 1}, \{M\}_{22}, \ldots, \{M\}_{r 2}, \ldots, \{M\}_{rr}]^T.$$ 

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

The distance between linear subspaces can be measured in terms of the canonical angles, formally defined as follows. Given two Stiefel matrices $U, U_0 \in O(p, r)$, let $\sigma_1(U_0^T U) \geq \ldots \geq \sigma_r(U_0^T U) \geq 0$ be the singular values of $U_0^T U$. Note the singular values of $U_0^T U$ are unitarily invariant and only depend on $\text{Span}(U)$ and $\text{Span}(U_0)$. The canonical angles between $U_0$ and $U$ are defined to be the diagonal entries of $\Theta(U_0, U) := \text{diag}(|\sin^{-1}\{\sigma_1(U_0^T U)\}|, \ldots, |\sin^{-1}\{\sigma_r(U_0^T U)\}|) \in \mathbb{R}^{r \times r}$. Then the spectral sine-theta distance and the Frobenius sine-theta distance between $\text{Span}(U_0)$ and $\text{Span}(U)$ are defined by $\|\sin \Theta(U_0, U)\|_2$ and $\|\sin \Theta(U_0, U)\|_F$, 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 \( S \subset \mathbb{R}^p \) is an \( r \)-dimensional linear subspace in \( \mathbb{R}^p \). It is always possible to find a Stiefel matrix \( U \in O(p, r) \) whose columns span \( S \), and one may use \( U \) as a representer for the subspace \( S \). The disadvantage of this representation is that \( U \) cannot be uniquely identified by \( S \), since for any orthogonal rotation matrix \( W \), \( \text{Span}(U) = \text{Span}(UW) \). 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 \( U = [Q_1^T, Q_2^T]^T \in O(p, r) \) such that \( Q_1 \in \mathbb{R}^{r \times r} \) is symmetric positive definite with respect to the uniform probability measure of \( \mathcal{G}(p, r) \). Formally, denote

\[
\mathcal{O}_+(p, r) = \left\{ U = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} \in O(p, r) : Q_1 \in M_+(r) \right\}
\]

and suppose \( l : \mathcal{O}_+(p, r) \to \mathcal{G}(p, r) \) is the map defined by \( l(U) = \text{Span}(U) \). 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 \( \mathcal{O}_+(p, r) \), and therefore, finding a suitable representation of subspaces reduces to finding a suitable representation of Stiefel matrices in \( \mathcal{O}_+(p, r) \). The case where the subspace \( S \) cannot be represented by a \( U \in \mathcal{O}_+(p, r) \) will be addressed using a standard patching argument in differential manifold theory (see Section 4.1 for details).

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

\[
\begin{align*}
\text{(2.1)} & \quad U : \varphi \in \mathbb{R}^{(p-r)r} \mapsto U(\varphi) := (I_p + X_\varphi)(I_p - X_\varphi)^{-1}I_{p \times r} \in \mathcal{O}(p, r), \\
\text{(2.2)} & \quad X_\varphi = \begin{bmatrix} \mathbb{R}^{r \times r} & -A^T \\ A & (p-r) \times (p-r) \end{bmatrix}, \quad \text{and} \quad I_{p \times r} = \begin{bmatrix} I_r \\ (p-r) \times r \end{bmatrix}.
\end{align*}
\]

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

\[
U = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} (I_r - A^T A)(I_r + A^T A)^{-1} \\ 2A(I_r + A^T A)^{-1} \end{bmatrix}.
\]

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

\[
\text{(2.3)} \quad DU(\varphi) = 2[I_{p \times r}^T(I_p - X_\varphi)^{-T} \otimes (I_p - X_\varphi)^{-1}] \Gamma_\varphi.
\]
where $\Gamma_\varphi$ is a matrix such that $\Gamma_\varphi \varphi = \text{vec}(X \varphi)$. An explicit formula for $\Gamma_\varphi$ is also available [34]: $\Gamma_\varphi = (I_p - Kp)(\Theta_1^T \otimes \Theta_2^T)$, where $\Theta_1 = \Gamma_{p \times r}$, and $\Theta_2 = [(p-r) \times r, I_{p-r}]$. 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 $U : \varphi \mapsto U(\varphi)$ be the Cayley parameterization defined by (2.1). Denote $DU(\varphi)$ the Fréchet derivative of $U$ given by (2.3). Then for all $\varphi, \varphi_0$, there exists a matrix $R_U(\varphi, \varphi_0) \in \mathbb{R}^{p \times r}$, such that

$$
U(\varphi) - U(\varphi_0) = 2(I_p - X_{\varphi_0})^{-1}(X_{\varphi} - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1}I_{p \times r} + R_U(\varphi, \varphi_0),
$$

where the remainder $R_U$ satisfies

$$
\|R_U(\varphi, \varphi_0)\|_F \leq 8\|\varphi - \varphi_0\|_F^2.
$$

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

**Proof.** First observe that by definition of $\Gamma_\varphi$, we have,

$$
\|\Gamma_\varphi\|_2 = \max_{\|\varphi\|_2 = 1} \|\Gamma_\varphi \varphi\|_2 = \max_{\|\varphi\|_2 = 1} \|\text{vec}(X_\varphi)\|_2 = \max_{\|\varphi\|_2 = 1} (2\|\text{vec}(A)\|_2^2)^{1/2} = \sqrt{2}.
$$

Because $X_\varphi^T \varphi = -X_\varphi$, we also have

$$
\|(I_p + X_\varphi)^{-1} - (I_p - X_{\varphi_0})^{-1}\|_2 = \|(I_p - X_\varphi)^{-1}(X_\varphi - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1}\|_2
$$

$$
= \lambda_{min}^{-1/2}(I_p + X_\varphi X_\varphi^T) \leq \lambda_{min}^{-1/2}(I_p) = 1
$$

for any $\varphi$. Write

$$
\|(I_p - X_{\varphi})^{-1} - (I_p - X_{\varphi_0})^{-1}\|_F = \|(I_p - X_{\varphi})^{-1}(X_{\varphi} - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1}\|_F
$$

$$
\leq \|(I_p - X_{\varphi})^{-1}\|_2\|X_{\varphi} - X_{\varphi_0}\|_F\|(I_p - X_{\varphi_0})^{-1}\|_2
$$

$$
\leq \|X_{\varphi} - X_{\varphi_0}\|_F.
$$

Furthermore, by matrix algebra,

$$
(I_p - X_{\varphi})^{-1} - (I_p - X_{\varphi_0})^{-1}
$$

$$
= (I_p - X_{\varphi})^{-1}(X_{\varphi} - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1}
$$

$$
= (I_p - X_{\varphi_0})^{-1}(X_{\varphi} - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1}
$$

$$
+ \{(I_p - X_{\varphi})^{-1} - (I_p - X_{\varphi_0})^{-1}\}(X_{\varphi} - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1}.
$$

Denote $T(\varphi, \varphi_0) := \{(I_p - X_{\varphi})^{-1} - (I_p - X_{\varphi_0})^{-1}\}(X_{\varphi} - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1}$. It follows that

$$
(I_p - X_{\varphi})^{-1} - (I_p - X_{\varphi_0})^{-1} = (I_p - X_{\varphi_0})^{-1}(X_{\varphi} - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1} + T(\varphi, \varphi_0),
$$

and

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

This manuscript is for review purposes only.
Therefore, we can write

\[
U(\varphi) - U(\varphi_0) = (I_p + X_{\varphi_0} + X_\varphi - X_{\varphi_0})(I_p - X_\varphi)^{-1} - (I_p - X_{\varphi_0})^{-1}I_p \times r \\
+ (X_\varphi - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1}I_p \times r \\
= (I_p + X_\varphi)(I_p - X_{\varphi_0})^{-1}(X_\varphi - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1}I_p \times r \\
+ (I_p + X_{\varphi_0})T(\varphi, \varphi_0)I_p \times r \\
+ (X_\varphi - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1} - (I_p - X_{\varphi_0})^{-1}I_p \times r \\
+ (I_p - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1}(X_\varphi - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1}I_p \times r \\
= 2(I_p - X_{\varphi_0})^{-1}(X_\varphi - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1}I_p \times r + R_U(\varphi, \varphi_0),
\]

where

\[
R_U(\varphi, \varphi_0) := (I_p + X_{\varphi_0})T(\varphi, \varphi_0)I_p \times r \\
+ (X_\varphi - X_{\varphi_0})(I_p - X_{\varphi_0})^{-1} - (I_p - X_{\varphi_0})^{-1}I_p \times r.
\]

Using the aforementioned results, we further compute

\[
\|R_U(\varphi, \varphi_0)\|_F \leq \|I_p + X_{\varphi_0}\|_2\|T(\varphi, \varphi_0)\|_F + \|X_\varphi - X_{\varphi_0}\|_F^2 \\
\leq (1 + 2\|A\|_2)\|X_\varphi - X_{\varphi_0}\|_F^2 + \|X_\varphi - X_{\varphi_0}\|_F^2 \\
\leq 8\|\varphi - \varphi_0\|_F^2.
\]

The proof is thus completed. \qed

The authors of [34] also showed that the Cayley parameterization is one-to-one, and we refer to the inverse map \( A : \mathbb{O}_+(p, r) \to \{ A \in \mathbb{R}^{(p-r) \times r} : \|A\|_2 < 1 \} \) as the inverse Cayley parameterization. Formally, for any \( U = [Q_1^T, Q_2^T]^T \in \mathbb{O}_+(p, r) \) where \( Q_1 \in \mathbb{M}_+(r) \), the inverse Cayley parameterization of \( U \) is given by

\[
(2.4) \quad A = A(U) = Q_2(I_r + Q_1)^{-1}.
\]

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

**Theorem 2.2.** Let \( A(\cdot) \) be the inverse Cayley parameterization defined by (2.4). Then for all \( U, U_0 \in \mathbb{O}_+(p, r) \), \( \|A(U) - A(U_0)\|_F \leq 2\|U - U_0\|_F \).

**Proof.** Observe that

\[
\|(I_r + Q_1)^{-1} - (I_r + Q_01)^{-1}\|_F \\
= \|(I_r + Q_1)^{-1}((I_r + Q_01) - (I_r + Q_1)))(I_r + Q_01)^{-1}\|_F \\
\leq \|(I_r + Q_1)^{-1}\|_2\|Q_1 - Q_01\|_F\|(I_r + Q_01)^{-1}\|_2 \\
\leq \|Q_1 - Q_01\|_F
\]

because \( I_r + Q_1 \succeq I_r \) and \( I_r + Q_01 \succeq I_r \). Therefore,

\[
\|A(U) - A(U_0)\|_F \\
\leq \|Q_2 - Q_{02}\|_F\|(I_r + Q_1)^{-1}\|_2 + \|Q_{02}\|_2\|(I_r + Q_1)^{-1} - (I_r + Q_01)^{-1}\|_F \\
\leq \|Q_2 - Q_{02}\|_F + \|Q_{1} - Q_{01}\|_F \\
\leq 2\|U - U_0\|_F.
\]

The proof is thus completed. \qed
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 $G(p, r)$ as follows. For any $\text{Span}(Y) \in G(p, r)$ with $Y \in \mathbb{R}^{p \times r}$, define
\[
\mathcal{U}_Y = \{\text{Span}(Z) : \det(Z^T Y) \neq 0, Z \in \mathbb{R}^{p \times r}\} \subset G(p, r)
\]
and the cross-section mapping is given by $B \in \mathbb{R}^{(p-r) \times r} \mapsto \text{Span}(Y + Y_B B)$, where $Y_B \in \mathbb{R}^{p \times (p-r)}$ spans the orthogonal complement of $Y$. 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:
\[
B \mapsto U(I_r + B^T B)^{-1/2} + U_Y B(I + B^T B)^{-1/2}.
\]

See, for example, [4, 8, 25, 46, 57]. When $U = P I_{p \times r}$ for some $p \times p$ permutation matrix $P$, the above parameterization coincides with those constructed in [8] and [31].

In this case, up to a permutation parameterization (2.5) has a one-to-one correspondence with the Cayley parameterization (2.1) through the relation
\[
B = 2A(I_r - A^T A)^{-1}.
\]

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 $\Sigma$ with $\text{rank}(\Sigma) = r \leq p$. Let $\Sigma$ yield the spectral decomposition $\Sigma = V \Lambda V^T$, where $V \in \mathcal{O}(p, r)$ is the Stiefel matrix of eigenvectors, and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_r)$ is the diagonal matrix of non-zero eigenvalues of $\Sigma$ with $|\lambda_1| \geq \ldots \geq |\lambda_r| > 0$. In scenarios where the eigenvalues $\lambda_1, \ldots, \lambda_r$ may include multiplicity, the eigenvector matrix $V$ may only be determined up to an orthogonal matrix in $\mathcal{O}(r)$. Also note by the aforementioned analysis, for almost every $\text{Span}(V) \in G(p, r)$ (with respect to the uniform probability measure over $G(p, r)$), there exists another Stiefel matrix $U \in \mathcal{O}_+(p, r)$ such that $\text{Span}(V) = \text{Span}(U)$. It follows that almost every $\Sigma$ with $\text{rank}(\Sigma) = r$ can be reparameterized as $\Sigma = U \Lambda U^T$ for a $p \times r$ Stiefel matrix $U \in \mathcal{O}_+(p, r)$ and a $r \times r$ symmetric matrix $M \in \mathcal{M}(r)$, where the term “almost every” is with respect to the product measure of the uniform probability measure over $\mathcal{O}(p, r)$ and the Lebesgue measure over $\mathcal{M}(r)$. Then by the result of Section 2.2, there exists a unique $A \in \mathbb{R}^{(p-r) \times r}$, $||A||_2 < 1$, and $\varphi = \text{vec}(A)$, such that $U = U(\varphi)$, where $U(\cdot)$ is the Cayley parameterization defined...
by (2.1). Let $\mu = \text{vech}(M)$. Conversely, the matrix $M$ can be viewed as a function of $\mu$, denoted generically by $M(\cdot) : \mathbb{R}^{r(r+1)/2} \to \mathbb{M}(r)$, as the inverse of the map $M \mapsto \mu = \text{vech}(M)$. Therefore, by denoting $\theta = [\varphi^T, \mu^T]^T$, we can represent almost every $\Sigma$ with rank($\Sigma$) = $r$ through the following matrix-valued function, which is generically denoted by $\Sigma(\cdot)$:

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

where

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

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

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

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 $\theta$ is convex and connected. In fact, given any $\theta_1 = [\text{vec}(A_1)^T, \mu_1^T]^T, \theta_2 = [\text{vec}(A_2)^T, \mu_2^T]^T \in \mathcal{D}(p, r)$, where $A_1, A_2 \in \mathbb{R}^{(p-r)\times r}, \mu_1, \mu_2 \in \mathbb{R}^{r(r+1)/2}$ with $\|A_1\|_2 < 1$ and $\|A_2\|_2 < 1$, we have, for any $\lambda \in (0, 1), (1 - \lambda)\theta_1 + \lambda \theta_2 = \begin{bmatrix} (1 - \lambda)\text{vec}(A_1) + \lambda \text{vec}(A_2) \\ (1 - \lambda)\mu_1 + \lambda \mu_2 \end{bmatrix} = \begin{bmatrix} (1 - \lambda)\mu_1 + \lambda \mu_2 \end{bmatrix}.$

Since $\| (1 - \lambda)A_1 + \lambda A_2 \|_2 \leq (1 - \lambda)\|A_1\|_2 + \lambda \|A_2\|_2 < (1 - \lambda) + \lambda = 1$, it follows that $(1 - \lambda)\theta_1 + \lambda \theta_2 \in \mathcal{D}(p, r)$, and hence, $\mathcal{D}(p, r)$ is convex and connected.

This paper is primarily interested in the intrinsic perturbation analysis between $\Sigma_0$ and $\Sigma$ with $\Sigma, \Sigma_0 \in \mathcal{H}(p, r)$, and $E := \Sigma - \Sigma_0$ is assumed to be comparatively smaller than $\Sigma_0$ in magnitude. In many statistical problems, $\Sigma_0$ is the referential matrix of interest, but only the perturbed version $\Sigma$ is accessible. To be more specific, $\Sigma$ typically plays the role of a function of the observed data, namely, an estimator for the unknown $\Sigma_0$. By the aforementioned analysis, $\Sigma$ and $\Sigma_0$ can be represented by some Euclidean vectors $\theta, \theta_0 \in \mathcal{D}(p, r)$, such that $\Sigma = \Sigma(\theta)$ and $\Sigma_0 = \Sigma(\theta_0)$. In turn, the problem of estimating the unobserved referential matrix $\Sigma_0$ reduces to estimating the Euclidean representer $\theta_0$ by an estimator $\theta$, and hence, the perturbation analysis of $\Sigma$ naturally translates to the problem of the perturbation analysis of $\theta$.

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

$$(2.9) \quad D_\varphi \Sigma(\theta) := (I_p^2 + K_{pp})\{U(\varphi)M(\mu) \otimes I_p\}DU(\varphi),$$

$$(2.10) \quad D_\mu \Sigma(\theta) := \{U(\varphi) \otimes U(\varphi)\}D_r,$$

$$(2.11) \quad D\Sigma(\theta) := [D_\varphi \Sigma(\theta) \quad D_\mu \Sigma(\theta)]$$

where $DU(\varphi)$ is the Fréchet derivative of the Cayley parameterization defined by (2.3). When $\theta$ takes value at the referential Euclidean vector $\theta_0 = [\varphi_0^T, \mu_0^T]^T$ such that $\Sigma_0 = \Sigma(\theta_0) = U(\varphi_0)M(\mu_0)U(\varphi_0)^T$, we simply write $U_0 = U(\varphi_0)$ and $M_0 = M(\mu_0)$.

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 $\mathcal{M}(p, r)$ to the perturbation of the corresponding representing Euclidean vectors $\theta$, $\theta_0$ through a first-order Taylor expansion device.

**Theorem 3.1.** Under the setup and notations of Sections 2.2 and 2.4, if $\|A_0\|_2 < 1$ and $\|A\|_2 < 1$, then there exists a $p \times p$ matrix $R(\theta, \theta_0)$ depending on $\theta, \theta_0$, such that

$$
\Sigma(\theta) - \Sigma_0 = 2(I_p - X_{\phi_0})^{-1}(X_{\phi} - X_{\phi_0})(I_p - X_{\phi_0})^{-T}\Sigma_0
$$

(3.1)

$$
- 2\Sigma_0(I_p - X_{\phi_0})^{-1}(X_{\phi} - X_{\phi_0})(I_p - X_{\phi_0})^{-T}
+ U_0\{M(\mu) - M_0\}U_0^T + R(\theta, \theta_0),
$$

where the remainder $R(\theta, \theta_0)$ satisfies $\|R(\theta, \theta_0)\|_F \leq 16(1 + \|M_0\|_2)\|\theta - \theta_0\|_2^2$.

**Proof.** First observe that the following matrix decomposition holds:

$$
\Sigma(\theta) - \Sigma(\theta_0)
$$

$$
= U_0(M - M_0)U_0^T + U_0M_0\{U(\varphi) - U_0\}^T + \{U(\varphi) - U_0\}M_0U_0^T + R_{\Sigma}(\theta, \theta_0),
$$

where the remainder

$$
R_{\Sigma}(\theta, \theta_0) = U(\varphi)(M - M_0)\{U(\varphi) - U_0\}^T + \{U(\varphi) - U_0\}M_0\{U(\varphi) - U_0\}^T
$$

$$
+ \{U(\varphi) - U_0\}M - M_0U_0^T
$$

satisfies

$$
\|R_{\Sigma}(\theta, \theta_0)\|_F \leq 2\|M - M_0\|_F\|U(\varphi) - U_0\|_F + \|M_0\|_2\|U(\varphi) - U_0\|_2^2
$$

$$
\leq \|M - M_0\|_2^2 + \|U(\varphi) - U_0\|_F^2 + \|M_0\|_2\|U(\varphi) - U_0\|_2^2
$$

$$
\leq 2\|\mu - \mu_0\|_2^2 + (1 + \|M_0\|_2)\|U(\varphi) - U_0\|_F^2.
$$

By Theorem 2.1,

$$
\|U(\varphi) - U_0\|_F \leq 2\sqrt{2}\|\varphi - \varphi_0\|_2
$$

for all $\varphi$ and $\varphi_0$. Therefore,

$$
\|R_{\Sigma}(\theta, \theta_0)\|_F \leq 8(1 + \|M_0\|_2)\|\theta - \theta_0\|_2^2.
$$

Furthermore, using Theorem 2.1 again, we obtain

$$
\text{vec}\{U(\varphi) - U_0\} = DU(\varphi_0)(\varphi - \varphi_0) + \text{vec}\{R_U(\varphi, \varphi_0)\},
$$

where $\|R_U(\varphi, \varphi_0)\|_F \leq 8\|\varphi - \varphi_0\|_2^2$ for all $\varphi, \varphi_0$. In matrix form, we have

$$
U(\varphi) - U_0 = 2(I_p - X_0)^{-1}(X_{\phi} - X_0)(I_p - X_0)^{-1}I_p^r + R_U(\varphi, \varphi_0).
$$

Hence we finally obtain

$$
\text{vec}\{\Sigma(\theta) - \Sigma_0\} = D_{\mu}\Sigma(\theta_0)(\mu - \mu_0) + (I_p^2 + K_{pp})(U_0M_0 \otimes I_p)\text{vec}\{U(\varphi) - U_0\}
$$

$$
+ \text{vec}\{R_{\Sigma}(\theta, \theta_0)\}
$$

$$
= D\Sigma(\theta_0)(\theta - \theta_0) + (I_p^2 + K_{pp})(U_0M_0 \otimes I_p)\text{vec}\{R_U(\varphi, \varphi_0)\}
$$

$$
+ \text{vec}\{R_{\Sigma}(\theta, \theta_0)\}
$$

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

This manuscript is for review purposes only.
Therefore, \( \lambda \) Euclidean representing vectors. Its proof is provided in Appendix.

Result, asserts that the reverse statement is true: the perturbation of the representing

Theorem 3.1 immediately implies that, locally at \( \theta_0 \), Euclidean vectors can be well controlled by the perturbation of the original matrices

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 \( \Sigma_0 \).

**Theorem 3.2.** Under the setup and notations of Sections 2.2 and 2.4, if \( \| A_0 \|_2 < 1 \), \( \| A \|_2 < 1 \), \( M_0 \) and \( M \) are positive definite, and then

\[
\| \Sigma(\theta) - \Sigma(\theta_0) \|_F \leq \frac{1 - \| A_0 \|_2^2}{4\sqrt{2}(1 + \| A_0 \|_2^2)}.
\]

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 \( \| A \|_2, \| A_0 \|_2 < 1 \), then

\[
\| U(\varphi) - U(\varphi_0) \|_F \leq \frac{2\| U(\varphi)U(\varphi)^T - U(\varphi_0)U(\varphi_0)^T \|_F}{\min\{\lambda_r(Q_{01}), \lambda_r(Q_1)\}},
\]

\[
\| \varphi - \varphi_0 \|_2 \leq \frac{4\| U(\varphi)U(\varphi)^T - U(\varphi_0)U(\varphi_0)^T \|_F}{\min\{\lambda_r(Q_{01}), \lambda_r(Q_1)\}}.
\]

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

\[
|\lambda_r(Q_1) - \lambda_r(Q_{01})| = \frac{|\lambda_r(Q_1) - \lambda_r(Q_{01})|}{\lambda_r(Q_1)} \leq \| Q_1 - Q_{01} \|_F \leq \frac{\| UU^T - U_0U_0^T \|_F}{\lambda_r(Q_{01})}.
\]

Note that \( \lambda_r(Q_{01}) = \lambda_r\{ (I_r - A_0^T A_0)(I_r + A_0^T A_0)^{-1} \} = (1 - \| A_0 \|_2^2)/(1 + \| A_0 \|_2^2) \).

By the Davis-Kahan theorem (see, e.g., Theorem 2 in [62]),

\[
|\lambda_r(Q_1) - \lambda_r(Q_{01})| \leq \frac{\| UU^T - U_0U_0^T \|_F}{\lambda_r(Q_{01})} \leq 2\sqrt{2} \left( 1 + \frac{\| A_0 \|_2^2}{1 + \| A_0 \|_2^2} \right) \frac{\| \Sigma(\theta) - \Sigma(\theta_0) \|_F}{\lambda_r(M_0)}.
\]

Therefore,

\[
\lambda_r(Q_1) \geq \frac{1 - \| A_0 \|_2^2}{1 + \| A_0 \|_2^2} \leq \frac{2\sqrt{2}(1 + \| A_0 \|_2^2)}{(1 - \| A_0 \|_2^2)\lambda_r(M_0)} \frac{\| \Sigma(\theta) - \Sigma(\theta_0) \|_F}{\lambda_r(M_0)} \frac{1 - \| A_0 \|_2^2}{2(1 + \| A_0 \|_2^2)} = \frac{\lambda_r(Q_{01})}{2}.
\]
Hence, by Lemma 3.3 and the Davis-Kahan theorem, we have

\[
\|U - U_0\|_F \leq \frac{4\|UU^T - U_0U_0^T\|_F}{\lambda_r(Q_{01})} \leq \frac{8\sqrt{2}\|\Sigma(\theta) - \Sigma(\theta_0)\|_F}{\lambda_r(Q_{01})\lambda_r(M_0)}
\]

\[= \frac{8\sqrt{2}(1 + \|A_0\|_2^2)}{\lambda_r(M_0)(1 - \|A_0\|_2^2)} \|\Sigma(\theta) - \Sigma(\theta_0)\|_F,
\]

\[
\|\varphi - \varphi_0\|_2 \leq \frac{8\|UU^T - U_0U_0^T\|_F}{\lambda_r(Q_{01})} \leq \frac{16\sqrt{2}\|\Sigma(\theta) - \Sigma(\theta_0)\|_F}{\lambda_r(Q_{01})\lambda_r(M_0)}
\]

\[= \frac{16\sqrt{2}(1 + \|A_0\|_2^2)}{\lambda_r(M_0)(1 - \|A_0\|_2^2)} \|\Sigma(\theta) - \Sigma(\theta_0)\|_F.
\]

For the matrix \( M \) and the vector \( \mu \), we have,

\[
\|\mu - \mu_0\|_2 \leq \|M - M_0\|_F = \|U^T \Sigma(\theta) U - U_0^T \Sigma(\theta_0) U_0\|_F
\]

\[
\leq \|U^T \{\Sigma(\theta) - \Sigma(\theta_0)\} U\|_F + \|U^T \Sigma_0(U - U_0)\|_F + \|(U - U_0)^T \Sigma_0 U_0\|_F
\]

\[
\leq \|\Sigma(\theta) - \Sigma(\theta_0)\|_F + 2\lambda_1(M_0)\|U - U_0\|_F
\]

\[
\leq \left\{1 + \frac{16\sqrt{2}\lambda_1(M_0)(1 + \|A_0\|_2^2)}{\lambda_r(M_0)(1 - \|A_0\|_2^2)} \right\} \|\Sigma(\theta) - \Sigma(\theta_0)\|_F.
\]

Therefore, we conclude that

\[
\|\theta - \theta_0\|_2 \leq \|\varphi - \varphi_0\|_2 + \|\mu - \mu_0\|_2
\]

\[
\leq \left[1 + \frac{16\sqrt{2}(1 + \lambda_1(M_0))(1 + \|A_0\|_2^2)}{\lambda_r(M_0)(1 - \|A_0\|_2^2)} \right] \|\Sigma(\theta) - \Sigma(\theta_0)\|_F.
\]

When \( M = M_0 = I_r \), Theorem 3.2, together with the fact that \( \|UU^T - VV^T\|_F = \sqrt{2}\|\sin(\Theta(U, V))\|_F \) for any \( U, V \in \mathcal{O}(p, r) \), directly leads to the following corollary regarding intrinsic perturbation of subspaces.

**Corollary 3.4.** Under the setup and notations of Section 2.2, if \( \|A_0\|_2 < 1 \), \( \|A\|_2 < 1 \), and

\[
\|\sin \Theta(U(\varphi), U(\varphi_0))\|_F \leq \frac{(1 - \|A_0\|_2^2)^2}{8(1 + \|A_0\|_2^2)^2},
\]

then

\[
\|\varphi - \varphi_0\|_2 \leq \sqrt{2} \left\{1 + \frac{32\sqrt{2}(1 + \|A_0\|_2^2)}{(1 - \|A_0\|_2^2)^2} \right\} \|\sin \Theta(U(\varphi), U(\varphi_0))\|_F.
\]

On the other hand, the following reverse inequality always holds for all \( \varphi \) and \( \varphi_0 \):

\[
\|\sin \Theta(U(\varphi), U(\varphi_0))\|_F \leq 4\|\varphi - \varphi_0\|_2.
\]

Corollary 3.4 suggests that, locally around \( \varphi_0 \), the Frobenius sine-theta distance between \( \text{Span}(U(\varphi)) \) and \( \text{Span}(U(\varphi_0)) \) is equivalent to the Euclidean distance between their representing Euclidean vectors \( \varphi \) and \( \varphi_0 \). Furthermore, by taking Theorem 2.1 into consideration, we conclude directly that \( \|\sin \Theta(U(\varphi), U(\varphi_0))\|_F \) is locally equiva- lent to \( \|U(\varphi) - U(\varphi_0)\|_F \). 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 \( \|A_0\|_2 < 1 \), and

\[
\| \sin \Theta \{ U(\varphi), U(\varphi_0) \} \|_F \leq \frac{(1 - \|A_0\|_2^2)^2}{8(1 + \|A_0\|_2^2)^2},
\]

then

\[
\| U(\varphi) - U(\varphi_0) \|_F \leq 4 \left\{ 1 + \frac{32 \sqrt{2}(1 + \|A_0\|_2^2)}{(1 - \|A_0\|_2^2)^2} \right\} \| \sin \Theta \{ U(\varphi), U(\varphi_0) \} \|_F.
\]

On the other hand, the following reverse inequality always holds for all \( \varphi \) and \( \varphi_0 \):

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

Given two Stiefel matrices \( V \) and \( V_0 \) in \( \mathcal{O}(p, r) \), their Frobenius sine-theta distance is equivalent to \( \| V - V_0 W^* \|_F \), where \( W^* \) solves the orthogonal Procrustes problem \( \min_{W \in \mathcal{O}(r)} \| V - V_0 W \|_F \) and can be computed explicitly using \( V_0 \) and \( V \). Formally, by Lemma 4.1 in [64] (also see Lemma 1 in [14]), we have

\[
\| \sin \Theta(V, V_0) \|_F \leq \| V - V_0 W^* \|_F \leq \sqrt{2} \| \sin \Theta(V, V_0) \|_F.
\]

As pointed out in [14], it is sometimes more convenient to work with the explicit expression \( \| V - V_0 W^* \|_F \) 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 \( W^* \) may still cause inconvenience for some theoretical analyses. In contrast, Theorem 3.5 loosely asserts that, in a small neighborhood of \( V_0 \) (or equivalently, a small neighborhood of \( \text{Span}(V_0) \) with respect to the sine-theta metric), by finding suitable representing Stiefel matrices \( U(\varphi), U(\varphi_0) \in \mathcal{O}_+(p, r) \) such that \( \text{Span}(U(\varphi)) = \text{Span}(V) \) and \( \text{Span}(U(\varphi_0)) = \text{Span}(V_0) \), \( \| \sin \Theta(V, V_0) \|_F \) is equivalent to \( \| U(\varphi) - U(\varphi_0) \|_F \), circumventing the orthogonal alignment matrix \( W^* \) and facilitating many technical analyses. Therefore, the metric \( \| U(\varphi) - U(\varphi_0) \|_F \) 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 \( \Sigma(\cdot) : \mathcal{O}(p, r) \to \mathcal{S}(p, r) \) 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 \( \sigma_r(M_0) \) and \( \|A_0\|_2 \).

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

\[
\sigma_{\min} \{ D \varphi \Sigma(\theta_0) \} \geq \begin{cases} 
\frac{2 \sqrt{2} \sigma_r(M_0)(1 - \|A_0\|_2^2)}{(1 + \|A_0\|_2^2)^2}, & \text{if } r \geq 2, \\
\frac{2 \sqrt{2} \sigma_r(M_0)}{1 + \|A_0\|_2^2}, & \text{if } r = 1.
\end{cases}
\]

Furthermore,

\[
\left\| \{ D \Sigma(\theta_0)^T D \Sigma(\theta_0) \}^{-1} \right\|_2 \leq \begin{cases} 
1 + \frac{(1 + 64 \|M_0\|_2^2)(1 + \|A_0\|_2^2)^4}{8 \lambda^2(M_0)(1 - \|A_0\|_2^2)^2}, & \text{if } r \geq 2, \\
1 + \frac{(1 + 64 \|M_0\|_2^2)(1 + \|A_0\|_2^2)^2}{8 \lambda^2(M_0)}, & \text{if } r = 1.
\end{cases}
\]
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,

\[
\Sigma_0^2 \otimes I_p - \Sigma_0 \otimes \Sigma_0 + I_p \otimes \Sigma_0^2 \geq \lambda^2(M_0) \{ U_0 U_0^T \otimes (I_p - U_0 U_0^T) + (I_p - U_0 U_0^T) \otimes U_0 U_0^T \},
\]

\[
\Sigma_0^2 \otimes (I_p - U_0 U_0^T) + (I_p - U_0 U_0^T) \otimes \Sigma_0^2 \geq \lambda^2(M_0) \{ U_0 U_0^T \otimes (I_p - U_0 U_0^T) + (I_p - U_0 U_0^T) \otimes U_0 U_0^T \}.
\]

**Lemma 3.8.** Let \( A_0 \in \mathbb{R}^{(p-r) \times r} \) with \( \|A_0\|_2 < 1 \), and define

\[
C_{11} = (I_r + A_0^T A_0)^{-1}, \quad C_{12} = -(I_r + A_0^T A_0)^{-1} A_0^T,
\]

\[
C_{21} = A_0 (I_r + A_0^T A_0)^{-1}, \quad C_{22} = I_p - r - A_0 (I_r + A_0^T A_0)^{-1} A_0^T.
\]

Let

\[
C_0 = \begin{bmatrix}
C_{11} & C_{12} \\
C_{21} & C_{22}
\end{bmatrix}, \quad \text{and} \quad U_0 = C_0^T C_0 I_{p \times r}.
\]

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

\[
\text{vec}(X_{\varphi}^T) (C_0^T \otimes C_0^T) \{ U_0 U_0^T \otimes (I_p - U_0 U_0^T) \} \{ C_0 \otimes C_0 \} \text{vec}(X_{\varphi})
\]

\[
= \text{vec}(X_{\varphi}^T) (C_0^T \otimes C_0^T) \{ (I_p - U_0 U_0^T) \otimes U_0 U_0^T \} \{ C_0 \otimes C_0 \} \text{vec}(X_{\varphi})
\]

\[
= \| C_{22} A C_{11} - C_{12} A^T C_{21} \|^2_F,
\]

where

\[
X_{\varphi} := \begin{bmatrix}
A & -A^T \\
I_r & (p-r) \times (p-r)
\end{bmatrix}.
\]

(ii) \( C_{11} \otimes C_{22} - (C_{21}^T \otimes C_{12}) K_{(p-r)r} \) has the following lower bound in spectra:

\[
\sigma_{\min}(C_{11} \otimes C_{22} - (C_{21}^T \otimes C_{12}) K_{(p-r)r}) \geq \begin{cases}
\frac{1 - \|A_0\|^2_2}{(1 + \|A_0\|^2_2)^2}, & \text{if } r > 1,
1 + \|A_0\|^2_2, & \text{if } r = 1.
\end{cases}
\]

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

\[
X_{\varphi} := \begin{bmatrix}
A & -A^T \\
I_r & (p-r) \times (p-r)
\end{bmatrix}.
\]

Write \( X_0 = X_{\varphi_0} \) and \( C_0 = (I_p - X_0)^{-1} \). Let \( A_0 \) be the corresponding \( (p-r) \times r \) matrix such that \( \varphi_0 := \text{vec}(A_0) \). Then for any \( \varphi = \text{vec}(A) \in \mathbb{R}^d \), we have \( \Gamma_{\varphi} \varphi = \text{vec}(X_{\varphi}) \) by the definition of \( \Gamma_{\varphi} \). Therefore,

\[
(U_0 M_0 \otimes I_r) D U(\varphi_0) \varphi = 2(U_0 M_0 \otimes I_r) [I_{p \times r}^T (I_p - X_0)^{-1} \otimes (I - X_0)^{-1}] \text{vec}(X_{\varphi})
\]

\[
= 2 [U_0 M_0 I_{p \times r}^T (I_p - X_0)^{-1} \otimes (I - X_0)^{-1}] \text{vec}(X_{\varphi})
\]

\[
= 2 \text{vec} \{ (I_p - X_0)^{-1} X_{\varphi} (I_p - X_0)^{-1} I_{p \times r} M_0 U_0^T \}
\]

\[
= 2 \text{vec} \{ (I_p - X_0)^{-1} X_{\varphi} (I_p - X_0)^{-1} U_0 M_0 U_0^T \}.
\]
By definition of the commutation matrix $K_{pp}$, $K_{pp}\vec{c}(M) = \vec{c}(M)$ for any symmetric $M \in \mathbb{R}^{p \times p}$ and $K_{pp}\vec{c}(M) = -\vec{c}(M)$ for any skew-symmetric $M \in \mathbb{R}^{p \times p}$.

Hence, for any $p \times p$ matrix $M$, we have, $(I_p^2 + K_{pp})\vec{c}(M) = \vec{c}(M + M^T)$, and hence,

$$D_r\Sigma(\theta_0) \varphi = (I_p^2 + K_{pp})(U_0 M_0 \otimes I_p) DU(\varphi_0) \varphi = 2(I_p^2 + K_{pp}) \vec{c}\{ (I_p - X_0)^{-1} X_0 (I_p - X_0)^{-1} U_0 M_0 U_0^T \} = 2 \vec{c}(C_0 X_\varphi C_0^T \Sigma_0 - \Sigma_0 C_0 X_\varphi C_0^T) = 2(\Sigma_0 C_0 \otimes C_0 - C_0 \otimes \Sigma_0 C_0) \vec{c}(X_\varphi).$$

Let $U_{0\perp}$ be the orthogonal complement of $U_0$ such that $[U_0, U_{0\perp}] \in \mathcal{O}(p)$. By Lemma 3.7,

$$
\|D_r \Sigma(\theta_0)\varphi\|_2^2 = 4 \vec{c}(X_\varphi)^T (\Sigma_0 C_0 \otimes C_0 - C_0 \otimes \Sigma_0 C_0)^T (\Sigma_0 C_0 \otimes C_0 - C_0 \otimes \Sigma_0 C_0) \vec{c}(X_\varphi) = 4 \vec{c}(X_\varphi)^T (C_0^T \otimes C_0^T)(C_0^T \otimes C_0 - C_0 \otimes C_0) \vec{c}(X_\varphi) = 4 \lambda^2_r(M_0) \vec{c}(X_\varphi)^T (C_0^T \otimes C_0^T) \vec{c}(X_\varphi) + 4 \lambda^2_r(M_0) \vec{c}(X_\varphi)^T (C_0^T \otimes C_0^T)(U_0 U_0^T \otimes U_0 U_0^T + U_0 U_0^T \otimes U_0 U_0^T)
$$

Write $C_0$ in the block form

$$C_0 = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} (I_r + A_0^T A_0)^{-1} & -(I_r + A_0^T A_0)^{-1} A_0^T \\ A_0 (I_r + A_0^T A_0)^{-1} & I_p - A_0 (I_r + A_0^T A_0)^{-1} A_0^T \end{bmatrix}$$

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

$$\|D_r \Sigma(\theta_0)\varphi\|_2^2 \geq 8 \lambda^2_r(M_0) \sigma_{\min}^2 \begin{bmatrix} (C_{11} \otimes C_{22}) - (C_{21}^T \otimes C_{12}^T) K_{(p-r)k}, & \vec{c}(A) \end{bmatrix} \|\varphi\|_2^2,$$

which immediately implies that

$$\sigma_{\min} \{ D_r \Sigma(\theta_0) \} \geq 2 \sqrt{2} \sigma_r(M_0) \sigma_{\min} \{ (C_{11} \otimes C_{22}) - (C_{21}^T \otimes C_{12}^T) K_{(p-r)k} \}.$$

By Lemma 3.8 (ii), we finally obtain

$$\sigma_{\min} \{ D_r \Sigma(\theta_0) \} \geq \begin{cases} \frac{2 \sqrt{2} \sigma_r(M_0) (1 - \|A_0\|_2^2)}{(1 + \|A_0\|_2^2)^2}, & \text{if } r > 1, \\ \frac{2 \sqrt{2} \sigma_r(M_0)}{1 + \|A_0\|_2^2}, & \text{if } r = 1. \end{cases}$$

completing the proof of the first assertion regarding $\sigma_{\min} \{ D_r \Sigma(\theta_0) \}$. ☐
Proof of second assertion of Theorem 3.6. Write

\[ D\Sigma(\theta_0)^T D\Sigma(\theta_0) = \begin{bmatrix} J_1 & J_2 & J_3 \end{bmatrix} := \begin{bmatrix} D_\varphi \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) & D_\varphi \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \\ D_\mu \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) & D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \end{bmatrix}. \]

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

\[ J_1 - J_2 J_3^{-1} J_2^T = D_\varphi \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) \]

\[ - D_\varphi \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \{ D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \}^{-1} D_\mu \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0), \]

We now assume that \( J_1 - J_2 J_3^{-1} J_2^T \) is invertible. Then by the block matrix inversion formula,

\[ \{ D\Sigma(\theta_0)^T D\Sigma(\theta_0) \}^{-1} = \begin{bmatrix} (J_1 - J_2 J_3^{-1} J_2^T)^{-1} & -(J_1 - J_2 J_3^{-1} J_2^T) J_2 J_3^{-1} \\ -J_3^{-1} (J_1 - J_2 J_3^{-1} J_2^T)^{-1} & J_3^{-1} + J_3^{-1} J_2^T (J_1 - J_2 J_3^{-1} J_2^T)^{-1} J_2 J_3^{-1} \end{bmatrix}, \]

By construction,

\[ \| J_3^{-1} \|_2 = \| \{ D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \}^{-1} \|_2 = \| \{ D_\mu^T (I_r \otimes I_r) D_\mu \}^{-1} \| \leq 1, \]

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

\[ \| J_2 \|_2 \leq \| D U (\varphi_0)^T (M_0 U_0^T \otimes I_p)(I_{p^2} + K_{pp})(U_0 \otimes U_0) D_r \| \leq 2 \| D U (\varphi_0) \|_2 \| M_0 \|_2 \| D_r \|_2 \]

\[ \leq 8 \| M_0 \|_2 \| I_{p \times r} (I_p - X_{\varphi_0})^{-1} \|_2 \| (I_p - X_{\varphi_0})^{-1} \|_2 \leq 8 \| M_0 \|_2. \]

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

\[ \| \{ D\Sigma(\theta_0)^T D\Sigma(\theta_0) \}^{-1} \|_2 \leq \| (J_1 - J_2 J_3^{-1} J_2^T)^{-1} \|_2 \]

\[ + \| J_3^{-1} + J_3^{-1} J_2^T (J_1 - J_2 J_3^{-1} J_2^T)^{-1} J_2 J_3^{-1} \|_2 \]

\[ \leq \| (J_1 - J_2 J_3^{-1} J_2^T)^{-1} \|_2 \]

\[ + \| J_3^{-1} \|_2 \| J_3^{-1} \|_2 \| J_2 \|_2 \| (J_1 - J_2 J_3^{-1} J_2^T)^{-1} \|_2 \]

\[ \leq 1 + (1 + 64 \| M_0 \|_2^2) \| (J_1 - J_2 J_3^{-1} J_2^T)^{-1} \|_2. \]

Therefore, it is sufficient to provide a lower bound for the smallest eigenvalue of

\[ J_1 - J_2 J_3^{-1} J_2^T. \]

For any \( \varphi \), we follow the computation above and write

\[ \varphi^T (J_1 - J_2 J_3^{-1} J_2^T) \varphi \]

\[ = \varphi^T D_\varphi \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) \varphi \]

\[ - \varphi^T D_\varphi \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \{ D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \}^{-1} D_\mu \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) \varphi \]

\[ = 4 \text{vec}(X_{\varphi})^T \{ (C_T^0 \otimes C_T^0)(\Sigma_0^0 \otimes I_p - 2 \Sigma_0 \otimes \Sigma_0 + I_p \otimes \Sigma_0^0)(C_0 \otimes C_0) \}^T \text{vec}(X_{\varphi}) \]

\[ - 4 \text{vec}(X_{\varphi})(C_T^0 \otimes C_T^0)(\Sigma_0 \otimes I_p - I_p \otimes \Sigma) \]

\[ \times D_\mu \Sigma(\theta_0) \{ D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \}^{-1} D_\mu \Sigma(\theta_0) \]

\[ \times (\Sigma_0 \otimes I_p - I_p \otimes \Sigma)(C_0 \otimes C_0) \text{vec}(X_{\varphi}). \]
Note that
\[
D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) (D_\mu \Sigma(\theta_0))^{-1} D_\mu \Sigma(\theta_0) = (U_0 \otimes U_0) D_r (I_r \otimes I_r) D_r (I_r \otimes I_r) D_r^{-1} D_r (U_0 \otimes U_0)^T
\]
is a projection matrix. Since \(D_r \in \mathbb{R}^{r^2 \times r(r+1)/2}\) has full column rank and \(r^2 \geq r(r+1)/2\), then
\[
D_r (I_r \otimes I_r) D_r^{-1} D_r \preceq I_r^2,
\]
and hence,
\[
D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) (D_\mu \Sigma(\theta_0))^{-1} D_\mu \Sigma(\theta_0) \preceq (U_0 \otimes U_0) (U_0 \otimes U_0)^T.
\]
Therefore, we invoke Lemmas 3.7 and 3.8 and proceed to compute
\[
\varphi^T (J_1 - J_2 J_3^{-1} J_2^T) \varphi \geq 4 \text{vec} (X_{\varphi})^T \{ (C_{0}^T \otimes C_{0}^T) (\Sigma_0^2 \otimes I_p - 2 \Sigma_0 \otimes \Sigma_0 + I_p \otimes \Sigma_0^2) (C_0 \otimes C_0) \} \text{vec} (X_{\varphi})
\]
\[
- 4 \text{vec} (X_{\varphi}) (C_{0}^T \otimes C_{0}^T) (\Sigma_0^2 \otimes I_p - I_p \otimes \Sigma_0) (U_0 \otimes U_0) (U_0 \otimes U_0)^T
\]
\[
\times (\Sigma_0 \otimes I_p - I_p \otimes \Sigma_0)
\]
\[
= 4 \text{vec} (X_{\varphi})^T \{ (C_{0}^T \otimes C_{0}^T) (\Sigma_0^2 \otimes I_p - 2 \Sigma_0 \otimes \Sigma_0 + I_p \otimes \Sigma_0^2) (C_0 \otimes C_0) \} \text{vec} (X_{\varphi})
\]
\[
- 4 \text{vec} (X_{\varphi}) (C_{0}^T \otimes C_{0}^T) (\Sigma_0^2 \otimes U_0 U_0^T - 2 \Sigma_0 \otimes \Sigma_0 + U_0 U_0^T \otimes \Sigma_0^2)
\]
\[
\times (C_0 \otimes C_0) \text{vec} (X_{\varphi})
\]
\[
= 4 \text{vec} (X_{\varphi})^T (C_{0}^T \otimes C_{0}^T) \{ \Sigma_0^2 \otimes (I_p - U_0 U_0^T) + (I_p - U_0 U_0^T) \otimes \Sigma_0^2 \}
\]
\[
\times (C_0 \otimes C_0) \text{vec} (X_{\varphi})
\]
\[
\geq 4 \lambda_2^2 (M_0) \text{vec} (X_{\varphi})^T (C_{0}^T \otimes C_{0}^T) (U_0 U_0^T \otimes U_0 U_0^T + U_0 U_0^T \otimes U_0 U_0^T)
\]
\[
\times (C_0 \otimes C_0) \text{vec} (X_{\varphi})
\]
\[
= 8 \lambda_2^2 (M_0) \| C_{22}^T A C_{11} - C_{21}^T A^T C_{21} \|_F^2
\]
\[
\geq 8 \lambda_2^2 (M_0) a_{\min}^2 \{ (C_{11} \otimes C_{22}) - (C_{21} \otimes C_{12}) K_{(p-r)r} \} \| \varphi \|_2^2.
\]
It follows from Lemma 3.8 (ii) that
\[
\lambda_{\min} (J_1 - J_2 J_3^{-1} J_2^T) \geq \begin{cases} 
8 \lambda_2^2 (M_0) (1 - \| A_0 \|_2^2) / (1 + \| A_0 \|_2^2)^2, & \text{if } r > 1, \\
8 \lambda_2^2 (M_0) / (1 + \| A_0 \|_2^2), & \text{if } r = 1.
\end{cases}
\]
Therefore, using (3.3),
\[
\| D \Sigma(\theta_0)^T D \Sigma(\theta_0) \|_2^{-1} \leq \begin{cases} 
1 + \frac{(1 + 64 \| M_0 \|_2^2) (1 + \| A_0 \|_2^2)^2}{8 \lambda_2^2 (M_0)} & \text{if } r > 1, \\
1 + \frac{(1 + 64 \| M_0 \|_2^2) (1 + \| A_0 \|_2^2)^2}{8 \lambda_2^2 (M_0)} & \text{if } r = 1,
\end{cases}
\]
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 $\mathcal{F}(p, r)$, 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 $P \in \mathcal{P}(p)$, define

$$x_P : \mathcal{D}(p, r) \rightarrow \mathcal{M}(p, r),$$

$$\theta \mapsto P\Sigma(\theta)P^T.$$

Then $\{(\mathcal{D}(p, r), x_P)\}_{P \in \mathcal{P}(p)}$ 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 Grassmannian structure of $\mathcal{G}(p, r)$ equivalent to that obtained in [3, 8, 26].

Every element $S \in \mathcal{G}(p, r)$ can be uniquely identified by a projection matrix $UU^T$ where $U \in \mathcal{O}(p, r)$ such that $S = \text{Span}(U)$. The topology on $\mathcal{G}(p, r)$ can be taken as the topology induced by the metric $\rho(U_1U_2^T, U_1^TU_2) = \|U_1U_2^T - U_1^TU_2\|_F$ for any $U_1, U_2 \in \mathcal{O}(p, r)$. 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 $\mathcal{A} = \{\text{vec}(A) : A \in \mathbb{R}^{(p-r) \times r}, \|A\|_2 < 1\}$. For each $P \in \mathcal{P}(p)$, define

$$y_P : \mathcal{A} \rightarrow \mathcal{G}(p, r),$$

$$\varphi = \text{vec}(A) \mapsto PU(\varphi)U(\varphi)^TP^T,$$

where $U(\cdot)$ is the Cayley parameterization given by (2.1). Then $\{\mathcal{A}, y_P\}_{P \in \mathcal{P}(p)}$ 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 \( \Xi \in T_B \mathcal{G}(p, r) \) has the form

\[
\Xi = Q \left[ r \times r \ Y^T \right] (p-r) \times (p-r) \left[ Y \right] Q^T,
\]

where \( \Pi = Q I_{p \times r} I_{p \times r}^T Q^T, Y \in \mathbb{R}^{(p-r) \times r} \), and for any \( \Xi_i \in T_B \mathcal{G}(p, r) \) with

\[
\Xi_i = Q \left[ r \times r \ Y_i^T \right] (p-r) \times (p-r) \left[ Y_i \right] Q^T, \quad i = 1, 2,
\]

the canonical metric is defined as \( \langle \Xi_1, \Xi_2 \rangle_{\Pi} = \text{tr}(Y_1^T Y_2) \) (see, for example, [3, 8, 26]).

By Lemma 3.8, for each \( i = 1, 2 \), there exists a unique matrix \( B_i \in \mathbb{R}^{(p-r) \times r} \) such that

\[
Y_i = C_{22} B_i C_{11} - C_{12}^T B_i^T C_{21}, \quad \text{where} \quad C_{11}, C_{12}, C_{21}, C_{22} \text{ are given by (3.2). Below,
}

we take advantage of such a relation between \( B_i \) and \( Y_i \) 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 \((A, y_P)\) is a parameterization of \( \mathcal{G}(p, r) \) at \( \Pi \) given by (4.1) for some \( p \times p \) permutation matrix \( P \), such that

\[
y_P(\text{vec}(A)) = \Pi, \quad \text{where vec}(A) \in A. \quad \text{Let} \quad \Xi_i \in T_B \mathcal{G}(p, r), \quad i = 1, 2, \quad \text{have representation}
\]

\[
\Xi_i = W \left[ C_{22} B_i C_{11} - C_{12}^T B_i^T C_{21} \right] (p-r) \times (p-r) \left[ C_{11} B_i^T C_{22} - C_{21}^T B_i C_{12} \right] W^T,
\]

where

\[
C_A = \left[ \begin{array}{cc} C_{11} & C_{12} \\ C_{21} & C_{22} \end{array} \right] = \left[ \begin{array}{cc} (I_r + A^T A)^{-1} & - (I_r + A^T A)^{-1} A^T \\ A(I_r + A^T A)^{-1} & I_{p-r} - A(I_r + A^T A)^{-1} A^T \end{array} \right],
\]

\( W = P C_A^{-T} C_A \), and \( B_i \in \mathbb{R}^{(p-r) \times r}, \quad i = 1, 2 \). Then the symmetric bilinear form

\[
\langle \Xi_1, \Xi_2 \rangle_{\Pi} := \langle C_{11} B_1 C_{22} - C_{12}^T B_1^T C_{21}, C_{11} B_2 C_{22} - C_{12}^T B_2^T C_{21} \rangle_{F}
\]

is positive definite, where \( \langle \cdot, \cdot \rangle_F \) is the Frobenius inner product between matrices.

It is clear that \( W \) defined in Proposition 4.2 plays the same role as \( Q \), and \( Y_i = C_{22} B_i C_{11} - C_{12}^T B_i^T C_{21} \) under the parameterization \((A, y_P)\) introduce in (4.1). This implies that the metric defined in Proposition 4.2 coincides with the canonical metric on \( \mathcal{G}(p, r) \). Consequentially, 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

\( \Pi = uu^T \) for some \( u \in \mathcal{O}(p, 1) \) and \( \Xi \in T_B \mathcal{G}(p, 1) \). 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

\[
\gamma(t) = \left\{ u \cos (\|u\|_2 t) + \frac{\Xi u}{\Xi u} \sin (\|u\|_2 t) \right\}^T \times \left\{ u \cos (\|u\|_2 t) + \frac{\Xi u}{\Xi u} \sin (\|u\|_2 t) \right\}^T,
\]

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.
For small \( t > 0 \), it can be shown that the geodesic distance between 
\( \gamma(0) = \Pi \) and \( \gamma(t) \) is given by \( d(\Pi, \gamma(t)) = \sigma t \), where \( \sigma := \|\Xi_u\|_2 \) is the canonical angle between \( u \) and \( u \cos \sigma t + (\Xi_u/\sigma) \sin \sigma t \) (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 \( G(p, 1) \) compared to several other distances. Suppose \( \Pi = uu^T \) with \( p = 200 \), \( u = p^{-1/2} 1_p \), \( \sigma = \|\Xi_u\|_2 = 1 \), and \( \Xi u = p^{-1/2} [1_{T_{p/2}}^T, -1_{T_{p/2}}^T]^T \), where \( 1_k \) is the vector of all ones in \( \mathbb{R}^k \) for \( k \geq 1 \). It is straightforward to show that \( \gamma(t) = y_1(\varphi(t)) \), where

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

Let \( u(t) = u \cos \sigma t + (\Xi_u/\sigma) \sin \sigma t \). We plot the geodesic distance \( d(u(t)u(t)^T, uu^T) \) on the Grassmannian \( G(p, 1) \), the Frobenius sine-theta distance \( \|\sin(\Theta(u(t), u))\|_F \) between subspaces, the Euclidean distance \( \|\varphi(t) - \varphi(0)\|_2 \) with regard to \( \varphi(t) \), and the Euclidean distance \( \|u(t) - u\|_2 \) with regard to \( u(t) \) 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.

**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 \( \Sigma, \Sigma_0 \in \mathcal{S}(p, r) \) and \( \Theta, \Theta_0 \in \mathcal{D}(p, r) \) such that \( \Sigma = \Sigma(\Theta) \) and \( \Sigma_0 = \Sigma(\Theta_0) \), where \( \mathcal{S}(p, r), \mathcal{D}(p, r), \) and \( \Sigma(\cdot) \) are defined in Section 2.4. The two matrices \( \Sigma \) and \( \Sigma_0 \) are related by a non-singular multiplicative factor \( D \in \mathbb{R}^{p\times p} \) simultaneously from the left and right, namely, \( \Sigma = D \Sigma_0 D^T \). The multiplicative perturbation analysis concerns the variation of the eigenspace of \( \Sigma \) relative to that of
Σ₀ as a function of Iᵣₚ − 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 Σ(·) be the matrix-valued function defined in (2.6) and \( θ, θ₀ ∈ \mathcal{D}(p, 1) \), where \( \mathcal{D}(p, 1) \) is defined in (2.7). Let \( θ = (φᵀ, µᵀ)ᵀ \) and \( θ₀ = (φ₀ᵀ, µ₀ᵀ)ᵀ \), where \( φ, φ₀ ∈ \mathbb{R}^{p⁻¹} \) with \( ∥φ∥_2, ∥φ₀∥_2 < 1 \). Suppose \( Σ(θ) = DΣ(θ₀)Dᵀ \) for some non-singular \( D ∈ \mathbb{R}^{p×p} \) and \( D − I_p \) has the following block form

\[
D − I_p = \begin{bmatrix} ε & αᵀ \\ β & Γ \end{bmatrix},
\]

where \( ε ∈ \mathbb{R}, α, β ∈ \mathbb{R}^{p⁻¹} \), and \( Γ ∈ \mathbb{R}^{(p⁻¹)×(p⁻¹)} \). If

\[
∥D − I_p∥_F ≤ \min \left\{ \begin{array}{c} 1 \\ \frac{4}{3} \end{array} \right\} \frac{1 − φ₀ᵀφ₀}{1 + φ₀ᵀφ₀},
\]

then

\[
∥φ − φ₀∥_2 ≤ \frac{(1 + φ₀ᵀφ₀)∥Du₀∥_2 − 1}{2 − 4∥D − I_p∥_F} + \frac{(1 − φ₀ᵀφ₀)(∥ε∥ + ∥β∥_2)}{2 − 4∥D − I_p∥_F} + \frac{2∥Γφ₀ − αᵀφ₀φ₀∥_2}{2 − 4∥D − I_p∥_F}.
\]

Proposition 4.4 implies that when the multiplicative matrix D has a small deviation Euclidean vectors \( ∥φ − φ₀∥_2 \) can be controlled by the deviation of D from I_p characterized in terms of the blocks \( ε, α, β, \) and \( Γ \) because \( ∥∥Du₀∥₂ − 1∥₂ ≤ ∥D − I_p∥_F \) and \( ∥Γφ₀ − αᵀφ₀φ₀∥_2 ≤ ∥Γ∥_₂ + ∥α∥_₂ \). 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 \( p = 200 \), \( Σ₀ = λ₀u₀u₀ᵀ, λ₀ = 10, u₀ = p⁻¹/₂I_p, \) and \( φ₀ = (p^{1/²} + 1)^⁻¹I_p⁻¹ \). Generate a multiplicative perturbation matrix \( Δ = [Δ_ij]_{p×p} \) with \( Δ_ij \) being independent and identically distributed N(0, 1) random variables and set \( D = I_p + cΔ/∥Δ∥_F \), where \( c = \min\{1/4, (1 − φ₀ᵀφ₀)/(4(1 + φ₀ᵀφ₀))\}⁻¹ \). Let \( u \) be the leading eigenvector of \( Σ := DΣ₀Dᵀ \). To find the representing vector \( φ ∈ \mathbb{R}^{p⁻¹} \) and \( µ ∈ \mathbb{R} \) such that \( Σ(θ) = DΣ₀Dᵀ \), where \( θ = (φᵀ, µᵀ)ᵀ \), note that

\[
DΣ₀Dᵀ = \left( \begin{bmatrix} Du₀ \\ ||Du₀||_2 \end{bmatrix} \right) \left( \begin{bmatrix} λ₀||Du₀||_2^₂ \\ ||Du₀||_2 \end{bmatrix} \right)ᵀ.
\]

From the proof of Proposition 4.4, we obtain that \( µ = λ₀||Du₀||_2^₂ \) and

\[
φ = [I_p⁻¹] \frac{Du₀}{||Du₀||_2} \left( 1 + \frac{I_p^{p×1}Du₀}{||Du₀||_2} \right)^⁻¹.
\]

In this case, one can construct \( Σ \) from \( θ = [φᵀ, µᵀ]ᵀ \) using the formula

\[
Σ = λ \left[ \begin{bmatrix} 1 − φᵀφ \\ 1 + φᵀφ \end{bmatrix} \begin{bmatrix} 1 − φᵀφ \\ 1 + φᵀφ \end{bmatrix} \right] \left( \begin{bmatrix} 2φᵀφ \\ 1 + φᵀφ \end{bmatrix} \right).
\]
In [39], the author showed the following relative perturbation bound:

$$\| \sin \Theta(u, u_0) \|_F \leq \sqrt{\| (I - D^{-1})u_0 \|_F^2 + \| (I - D)u_0 \|_F^2}.$$  

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 $$\| \sin \Theta(u, u_0) \|_2$$ between subspaces, the relative perturbation bound of [39], and the classical Davis-Kahan perturbation bound 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 $$\Sigma_0 \mapsto D\Sigma_0 D^T$$ and the block structure of the multiplicative factor $$D$$.

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 $$\Sigma$$ is a $$p_1 \times p_2$$ matrix with rank $$r$$ and let $$\Sigma = V_1 \Lambda V_2^T$$ be its singular value decomposition, where $$V_1 \in \mathcal{O}(p_1, r)$$, $$V_2 \in \mathcal{O}(p_2, r)$$, and $$\Lambda$$ is the diagonal matrix of the singular values of $$\Sigma$$. Assume that $$\Gamma_{p_2 \times r}^T V_2$$ is invertible and let $$\Gamma_{p_1 \times r}^T V_2 = W_1 \text{diag} \{ \sigma_1(\Gamma_{p_1 \times r}^T V_2), \ldots, \sigma_r(\Gamma_{p_1 \times r}^T V_2) \} W_2^T$$ be its singular value decomposition, where $$W_1, W_2 \in \mathcal{O}(r)$$. Similar to Section 2.4, we can parameterize the rectangular matrix $$\Sigma$$ using the two matrices $$U = V_2 W_2 W_1^T$$ and $$M = V_1 \Lambda W_2 W_1^T$$.  

This manuscript is for review purposes only.
Clearly, $\Sigma = MU^T$, where $M \in \mathbb{R}^{p_1 \times r}$ is a constraint-free full-rank matrix and $U \in \mathbb{O}_r(p_2, r)$. Invoking the Cayley parameterization (2.1), we can further reparameterize $U$ using a $(p_2 - r) \times r$ matrix $A$ with $\|A\|_2 < 1$, such that $U = U(\varphi)$, where $\varphi = \text{vec}(A) \in \mathbb{R}^{(p_2 - r)r}$. In particular, we see that $\text{Span}(V_2) = \text{Span}(U)$. Now denote $\theta = [\varphi^T, \mu^T]^T$, where $\mu = \text{vec}(M)$. Then we can view $\Sigma$ generically as a matrix-valued function of $\theta$:

\begin{equation}
\theta \mapsto \Sigma(\theta) = M(\mu)U(\varphi)^T,
\end{equation}

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

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 $\Sigma_0$ be $V_2$. We further assume $\Sigma_0$ is non-singular. Similarly, $\Sigma_0$ can also be represented by a Euclidean vector $\theta_0$. Write $\theta_0 = [\varphi_0^T, \mu_0^T]^T \in \mathbb{R}^{(p_2 - r)r} \times \mathbb{R}^{p_1 r}$, where $\varphi_0 = \text{vec}(A_0)$ for a $(p_2 - r) \times r$ matrix $A_0$ with $\|A_0\|_2 < 1$ and $\mu_0 = \text{vec}(M_0)$ for a $p_1 \times r$ full-rank matrix $M_0$. Let $U_0 := U(\varphi_0)$. We then define the following matrix-valued functions, extending the functions in (2.9) to general rectangular matrices:

\begin{align}
D_\varphi \Sigma(\theta) &= K_{p_2 p_1} (M \otimes I_{p_2}) DU(\varphi), \\
D_\mu \Sigma(\theta) &= U(\varphi) \otimes I_{p_1}, \\
D\Sigma(\theta) &= \begin{bmatrix} D_\varphi \Sigma(\theta) & D_\mu \Sigma(\theta) \end{bmatrix}.
\end{align}

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 Frèchet 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 $\|A_0\|_2 < 1$ and $\|A\|_2 < 1$, then there exists a $p_1 \times p_2$ matrix $R(\theta, \theta_0)$ depending on $\theta, \theta_0$, such that

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

where $R(\theta, \theta_0)$ satisfies $\|R(\theta, \theta_0)\|_F \leq (4 + 8\|M_0\|_2)\|\theta - \theta_0\|_F^2$ for all $\theta, \theta_0$. In matrix form, the above displayed equation can be written as

\[ \Sigma(\theta) - \Sigma_0 = 2M_0 U_0 (I_{p_2} - X_{\varphi_0})^{-1}(X_{\varphi} - X_{\varphi_0})(I_{p_2} - X_{\varphi_0})^{-T} + (M - M_0) U_0^T + R(\theta, \theta_0). \]

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

\[ \Sigma - \Sigma_0 = \Sigma(\theta) - \Sigma(\theta_0) = (M - M_0) U_0^T + M_0\{U(\varphi) - U_0\}^T + R(\Sigma, \theta_0), \]

where the remainder $R(\Sigma, \theta_0) = (M - M_0)\{U(\varphi) - U_0\}^T$ satisfies

\[ \|R(\Sigma, \theta_0)\|_F \leq \|M - M_0\|_F\|U(\varphi) - U_0\|_F \leq \frac{1}{2}\|\mu - \mu_0\|_2^2 + \frac{1}{2}\|U(\varphi) - U_0\|_F^2. \]

By Theorem 2.1, $\|U(\varphi) - U_0\|_F \leq 2\sqrt{2}\|\varphi - \varphi_0\|_2$. Hence, $\|R(\Sigma, \theta_0)\|_F \leq 4\|\theta - \theta_0\|_F^2$. Furthermore, using Theorem 2.1 again, we obtain

\[ \text{vec}\{U(\varphi) - U_0\} = DU(\varphi_0)(\varphi - \varphi_0) + \text{vec}\{R_U(\varphi, \varphi_0)\}, \]

This manuscript is for review purposes only.
where \(\|R_U(\varphi, \varphi_0)\|_F \leq 8\|\varphi - \varphi_0\|_2^2\) for all \(\varphi, \varphi_0\). In matrix form, we have
\[
U(\varphi) - U_0 = 2(I_p - X_0)^{-1}(X_\varphi - X_0)(I_p - X_0)^{-1}I_{p \times r} + R_U(\varphi, \varphi_0).
\]

Hence we finally obtain
\[
\text{vec}\{\Sigma(\theta) - \Sigma_0\} = D_{\mu} \Sigma(\theta_0)(\mu - \mu_0) + K_{p_2p_1}(M_0 \otimes I_{p_2})\text{vec}\{U(\varphi) - U_0\}
\]
\[
+ \text{vec}\{R_{\Sigma}(\theta, \theta_0)\} = D_{\Sigma}(\theta_0)(\theta - \theta_0) + \text{vec}\{R(\theta, \theta_0)\},
\]
where \(\|R(\theta, \theta_0)\|_F \leq \|M_0\|_2\|R_U(\varphi, \varphi_0)\|_F + \|R_{\Sigma}(\theta, \theta_0)\|_F \leq (4 + 8\|M_0\|_2)\|\theta - \theta_0\|_2^2\).

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

**Theorem 5.2.** Under the setup and notations of Sections 2.2 and 5,
\[
\|\{D_{\Sigma}(\theta_0)^T D_{\Sigma}(\theta_0)\}^{-1}\|_2 \leq \begin{cases} 
1 + \frac{(1 + 8\|M_0\|_2^2)(1 + \|A_0\|_2^2)^4}{4\sigma_f^2(M_0)(1 - \|A_0\|_2^2)^2}, & \text{if } r \geq 2, \\
1 + \frac{(1 + 8\|M_0\|_2^2)(1 + \|A_0\|_2^2)^2}{4\sigma_f^2(M_0)}, & \text{if } r = 1.
\end{cases}
\]

**Proof.** Let \(\theta_0 = [\varphi_0^T, \mu_0^T]^T\), where \(\varphi_0 = \text{vec}(A_0)\) for some \(A_0 \in \mathbb{R}^{(p_2 - r) \times r}\) and \(\mu_0 = \text{vec}(M_0)\) for some \(M_0 \in \mathbb{R}^{p_1 \times r}\). Let \(U_0 := U(\varphi_0)\). Then
\[
D_{\Sigma}(\theta_0)^T D_{\Sigma}(\theta_0) = \begin{bmatrix} J_1 & J_2 \\ J_2^T & J_3 \end{bmatrix} := \begin{bmatrix} D_\varphi \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) & D_\varphi \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \\ D_\mu \Sigma(\theta_0)^T D_\varphi \Sigma(\theta_0) & D_\mu \Sigma(\theta_0)^T D_\mu \Sigma(\theta_0) \end{bmatrix}
\]
\[
= \begin{bmatrix} DU(\varphi)^T (M_0^T M_0 \otimes I_{p_2}) DU(\varphi) & DU(\varphi)^T (M_0^T I_{p_2}) K_{p_2p_1}(U_0 \otimes I_{p_1}) \\ (U_0^T \otimes I_{p_1}) K_{p_2p_1}(M_0 \otimes I_{p_2}) DU(\varphi) & I_r \otimes I_{p_1} \end{bmatrix}.
\]

The Schur complement of \(J_3\) of the entire matrix \(D_{\Sigma}(\theta_0)^T D_{\Sigma}(\theta_0)\) is given by
\[
J_1 - J_2 J_3^{-1} J_2^T = DU(\varphi_0)^T (M_0^T M_0 \otimes I_{p_2}) DU(\varphi_0) - DU(\varphi_0)^T (M_0^T \otimes I_{p_2}) K_{p_2p_1}(U_0 U_0^T \otimes I_{p_1}) K_{p_2p_1}(M_0 \otimes I_{p_2}) DU(\varphi_0).
\]

Similar to the proof of Theorem 3.6, we provide a lower bound for the smallest eigenvalue of \(J_1 - J_2 J_3^{-1} J_2^T\). Denote
\[
X_0 = X_{\varphi_0} = \begin{bmatrix} r \times r & -A_0^T \\ A_0 & (I_{p_2} - X_{\varphi_0})^{-1} \end{bmatrix}, \quad C_0 = (I_{p_2} - X_{\varphi_0})^{-1}, \quad W_0 = C_0^{-T} C_0.
\]
For any nonzero vector \( \varphi \in \mathbb{R}^{(p_2-r)r} \), write
\[
\varphi^T(J_1 - J_2J_3^{-1}J_2^T)\varphi
= \varphi^TDU(\varphi)^T(M_2^T M_0 \otimes I_{p_2})DU(\varphi)\varphi
- \{K_{p_2p_1}(M_0 \otimes I_{p_2})DU(\varphi_0)\varphi\}^T(U_0U_0^T \otimes I_{p_1})\{K_{p_2p_1}(M_0 \otimes I_{p_2})DU(\varphi_0)\varphi\}
= 4\{(I_{p_2 \times r}C_0^T \otimes C_0)\text{vec}(X_\varphi)^T(M_2^T M_0 \otimes I_{p_2})\{I_{p_2 \times r}C_0^T \otimes C_0\}\text{vec}(X_\varphi)\}
- 4\{K_{p_2p_1}(M_0 \otimes I_{p_2})(I_{p_2 \times r}C_0^T \otimes C_0)\text{vec}(X_\varphi)\}^T(U_0U_0^T \otimes I_{p_1})
\times \{K_{p_2p_1}(M_0 \otimes I_{p_2})(I_{p_2 \times r}C_0^T \otimes C_0)\text{vec}(X_\varphi)\}
= 4\{\text{vec}(C_0 X_\varphi C_0 I_{p_2 \times r})\}^T(M_2^T M_0 \otimes I_{p_2})\{\text{vec}(C_0 X_\varphi C_0 I_{p_2 \times r})\}
- 4\{\text{vec}(C_0 X_\varphi C_0 I_{p_2 \times r})\}^T(M_2^T M_0 \otimes U_0U_0^T)\{\text{vec}(C_0 X_\varphi C_0 I_{p_2 \times r})\}
\times \{\text{vec}(C_0 X_\varphi C_0 I_{p_2 \times r})\}
= 4\{\text{vec}(C_0 X_\varphi C_0 I_{p_2 \times r})\}^T\{M_2^T M_0 \otimes (I_{p_2} - U_0U_0^T)\}\{\text{vec}(C_0 X_\varphi C_0 I_{p_2 \times r})\}
= 4\text{vec}(X_\varphi)(C_0^T \otimes C_0)\{U_0M_2^T M_0U_0^T \otimes (I_{p_2} - U_0U_0^T)\}\{\text{vec}(C_0 X_\varphi C_0 I_{p_2 \times r})\}.
\]

Let \( M_2^T M_0 = VSV^T \) be the spectral decomposition of \( M_2^T M_0 \), where \( V \in \mathbb{O}(r) \), and \( S \) is the diagonal matrix of the eigenvalues of \( M_2^T M_0 \). Let \( \tilde{V} = \text{diag}(V, I_{p_2-r}) \). Then we compute
\[
U_0M_2^T M_0U_0^T \otimes (I_{p_2} - U_0U_0^T)
= W_0 \tilde{V}\left[S_{(p_2-r) \times (p_2-r)} \right] \tilde{V}^T W_0^T \otimes W_0 \tilde{V}^{r \times r} I_{p_2-r} \tilde{V}^T W_0^T
= (W_0 \otimes W_0)(\tilde{V} \otimes \tilde{V})\left[S \otimes I_{p_2-r} \right] (p_2-r)^2 \times (p_2-r)^2
\times (\tilde{V} \otimes \tilde{V})^T(W_0 \otimes W_0)
\geq \sigma_r^2(M_0)(W_0 \otimes W_0)\left[V \otimes I_{p_2-r} \otimes \tilde{V}\right]
\times \left[I_r \otimes I_{p_2-r} \right] (p_2-r)^2 \times (p_2-r)^2
\times (W_0^T \otimes W_0^T)
= \sigma_r^2(M_0)(I_{(p_2-r) \times (p_2-r)\}) \otimes I_{p_2-r} \langle W_0^T \otimes W_0^T \rangle
= \sigma_r^2(M_0)\{U_0U_0 \otimes (I_{p_2} - U_0U_0^T)\}.$$
Therefore, the quadratic form $\varphi^T (J_1 - J_2 J_3^{-1} J_2^T) \varphi$ can be lower bounded:

$$\varphi^T (J_1 - J_2 J_3^{-1} J_2^T) \varphi \geq 4\sigma_r^2(M_0) \left\| \left( T_{p_2 \times r}^T C_0^T C_0^{-1} \right) \otimes \left( T_{(p_2-r) \times r}^T I_{p_2-r} \right) \right\|_F^2$$

$$= 4\sigma_r^2(M_0) \left\| \left( T_{p_2 \times r}^T C_0^T \right) \otimes \left( T_{(p_2-r) \times r}^T I_{p_2-r} \right) \right\|_F^2.$$

This follows that

$$\varphi^T (J_1 - J_2 J_3^{-1} J_2^T) \varphi \geq 4\sigma_r^2(M_0) \left\| \left( T_{p_2 \times r}^T C_0^T \right) \right\|_F^2$$

$$\geq 4\sigma_r^2(M_0) \left\| \left( T_{p_2 \times r}^T C_0^T \right) \right\|_F^2.$$

We finally invoke Lemma 3.8 to conclude that

$$\varphi^T (J_1 - J_2 J_3^{-1} J_2^T) \varphi \geq \begin{cases} 
\frac{4\sigma_r^2(M_0)}{(1 + \|A_0\|^2)^4} \varphi^2, & \text{if } r \geq 2, \\
\frac{4\sigma_r^2(M_0)}{(1 + \|A_0\|^2)^2} \varphi^2, & \text{if } r = 1,
\end{cases}$$

which further implies that

$$\| (J_1 - J_2 J_3^{-1} J_2^T)^{-1} \|_2 \leq \begin{cases} 
\frac{(1 + \|A_0\|^2)^4}{4\sigma_r^2(M_0)(1 - \|A_0\|^2)^2}, & \text{if } r \geq 2, \\
\frac{(1 + \|A_0\|^2)^2}{4\sigma_r^2(M_0)}, & \text{if } r = 1.
\end{cases}$$

This shows that $J_1 - J_2 J_3^{-1} J_2^T$ is invertible. Since $J_3$ is also invertible, the property of the Schur complement immediately implies that $D\Sigma(\theta_0)^T D\Sigma(\theta_0)$ is invertible.

Furthermore, by the block matrix inversion formula, we have

$$\left\{ D\Sigma(\theta_0)^T D\Sigma(\theta_0) \right\}^{-1} = \left[ \begin{array}{cc}
(\text{Id} - J_2 J_3^{-1} J_2^T)^{-1} & -(\text{Id} - J_2 J_3^{-1} J_2^T)^{-1} J_2 J_3^{-1} \\
-J_3^{-1} J_2 J_3^{-1} (\text{Id} - J_2 J_3^{-1} J_2^T)^{-1} J_2 J_3^{-1} & \text{Id} - J_3^{-1} J_2 J_3^{-1} (\text{Id} - J_2 J_3^{-1} J_2^T)^{-1} J_2 J_3^{-1}
\end{array} \right].$$

By construction, $\|J_3\|_2 = 1$ and

$$\|J_2\|_2 \leq \|DU(\varphi_0)^T (M_0^T \otimes I_p) K_{p_2 \times p_1} (U_0 \otimes I_{p_1})\| \leq \|DU(\varphi_0)\|_2 \|M_0\|_2 \leq 2\sqrt{2} \|M_0\|_2 \|I_p \otimes (X_{\varphi_0})^{-1}\|_2 \leq 2\sqrt{2} \|M_0\|_2.$$
Thus, by Lemma 3.4 of [10], we see that
\[ \| \{ D \Sigma(\theta_0)^T D \Sigma(\theta_0) \}^{-1} \|_2 \leq \| (J_1 - J_2 J_3^{-1} J_2^T)^{-1} \|_2 \]
\[ + \| J_3^{-1} + J_3^{-1} J_2^T (J_1 - J_2 J_3^{-1} J_2^T)^{-1} J_3 J_3^{-1} \|_2 \]
\[ \leq \| (J_1 - J_2 J_3^{-1} J_2^T)^{-1} \|_2 \]
\[ + \| J_3^{-1} \|_2 + \| J_3^{-1} \|_2 \| J_2 \|_2 \| (J_1 - J_2 J_3^{-1} J_2^T)^{-1} \|_2 \]
\[ \leq 1 + (1 + 8 \| M_0 \|_2^2) \| (J_1 - J_2 J_3^{-1} J_2^T)^{-1} \|_2. \]

The proof is completed by combining the upper bound for \( \| (J_1 - J_2 J_3^{-1} J_2^T)^{-1} \|_2 \). \( \square \)

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 \( \Sigma_0 \) may correspond to the ground truth of the parameter of interest, and \( \Sigma \) 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 \( U \in \mathbb{C}(p, r) \) has the maximum row-wise Euclidean norm that can be upper bounded by a constant factor of \( \sqrt{r/p} \).

Under the Cayley parameterization \( \varphi = \text{vec}(A) \mapsto U = U(\varphi) \), this requires that \( \| A \|_2 \) is close to 1, and the resulting Jacobian matrix \( DU(\varphi) \) 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 \( U = U(\varphi) \) and \( U_0 = U(\varphi_0) \).

First note that
\[ \| UU^T - U_0 U_0^T \|_F^2 = \left\| \begin{bmatrix} Q_1^2 - Q_{01}^2 & Q_1 Q_2^T - Q_{01} Q_{02}^T \\ Q_2 Q_1 - Q_{02} Q_{01} & Q_2 Q_2^T - Q_{02} Q_{02}^T \end{bmatrix} \right\|_F^2 \]
\[ = \| Q_1^2 - Q_{01}^2 \|_F^2 + 2 \| Q_2 Q_1 - Q_{02} Q_{01} \|_F^2 + \| Q_2 Q_2^T - Q_{02} Q_{02}^T \|_F^2. \]
Therefore, by Theorem 2.2, \( \| \varphi - \varphi_0 \|_2 = \| A - A_0 \|_F \leq 2 \| U - U_0 \|_F \). The proof is completed by combining the obtained upper bound for \( \| U - U_0 \|_F \).
Proof of Lemma 3.7. Let $U_{0,\perp}$ be the orthogonal complement of $U_0$ such that $W_0 := [U_0, U_{0,\perp}] \in \mathcal{O}(p)$, and denote

$$\tilde{M}_0 = I_{p \times r} M_0 I_{p \times r}^T = \begin{bmatrix} M_0 \\ (p-r) \times (p-r) \end{bmatrix}.$$  

Clearly, one can take $W_0 = C_0^{-T} C_0$. Suppose $M_0 = V_0 A_0 V_0^T$ is the spectral decomposition of $M_0$. Then $\tilde{M}_0 = I_{p \times r} V_0 A_0 V_0^T I_{p \times r} = \tilde{V}_0 \tilde{A}_0 \tilde{V}_0^T$, where

$$\tilde{V}_0 := \begin{bmatrix} V_0 \\ I_{p-r} \end{bmatrix}, \quad \text{and} \quad \tilde{A}_0 = \begin{bmatrix} A_0 \\ (p-r) \times (p-r) \end{bmatrix}.$$  

For convenience let $\lambda_{0k} := \lambda_k(M_0)$, $k \in [r]$. For the first assertion, write

$$\begin{aligned}
\Sigma_0^2 \otimes I_p - 2\Sigma_0 \otimes \Sigma_0 + I_p \otimes \Sigma_0^2 \\
= (W_0 \otimes W_0)(\tilde{M}_0 \otimes I_p - 2\tilde{M}_0 \otimes \tilde{M}_0 + I_p \otimes \tilde{M}_0^2)(W_0^T \otimes W_0^T) \\
= (W_0 \otimes W_0) \begin{bmatrix} \tilde{M}_0^2 \otimes I_p - 2\tilde{M}_0 \otimes \tilde{M}_0 + I_p \otimes \tilde{M}_0^2 \\ I_{(p-r) \times (p-r)} \otimes \tilde{M}_0^2 \end{bmatrix}(W_0^T \otimes W_0^T) \\
= (W_0 \otimes W_0) (\tilde{V}_0 \otimes \tilde{V}_0) \begin{bmatrix} A_0^2 \otimes I_p - 2A_0 \otimes \tilde{A}_0 + I_p \otimes \tilde{A}_0^2 \\ I_{(p-r) \times \tilde{A}_0^2} \end{bmatrix} \\
\times (\tilde{V}_0^T \otimes \tilde{V}_0^T)(W_0^T \otimes W_0^T) \\
\times (\tilde{V}_0^T \otimes \tilde{V}_0^T)(W_0^T \otimes W_0^T). 
\end{aligned}$$  

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

$$\begin{aligned}
(\lambda_{0k} I_p - \tilde{A}_0)^2 &\geq \lambda_{0r}^2 \begin{bmatrix} r \times r \\ I_{p-r} \end{bmatrix} \\
\text{and} \\
\tilde{A}_0^2 &\geq \lambda_{0r}^2 \begin{bmatrix} I_{r} \\ (p-r) \times (p-r) \end{bmatrix},
\end{aligned}$$

we can further write

$$\begin{aligned}
\Sigma_0^2 \otimes I_p - 2\Sigma_0 \otimes \Sigma_0 + I_p \otimes \Sigma_0^2 \\
\geq \lambda_{0r}^2 (W_0 \otimes W_0)(\tilde{V}_0 \otimes \tilde{V}_0) \begin{bmatrix} I_r \otimes \begin{bmatrix} r \times r \\ I_{p-r} \end{bmatrix} \\ I_{p-r} \otimes \begin{bmatrix} I_{r} \\ (p-r) \times (p-r) \end{bmatrix} \end{bmatrix} \times (\tilde{V}_0^T \otimes \tilde{V}_0^T)(W_0^T \otimes W_0^T). 
\end{aligned}$$
Now we focus on the matrix on the right-hand side of the previous display. Write

\[
\begin{align*}
\sum &\otimes \mathbf{I}_{p} - 2\sum \otimes \sum + \mathbf{I}_{p} \otimes \sum^{2} \\
&\geq \lambda_{pr}(\mathbf{W}_{0} \otimes \mathbf{W}_{0})(\tilde{\mathbf{V}}_{0} \otimes \tilde{\mathbf{V}}_{0}) \begin{bmatrix} \mathbf{I}_{r} \otimes [r \times r] & \mathbf{I}_{p-r} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{r} \\ \mathbf{I}_{p-r} \otimes \mathbf{I}_{r} \end{bmatrix} \\
&\times (\tilde{\mathbf{V}}_{0}^{T} \otimes \tilde{\mathbf{V}}_{0}^{T})(\mathbf{W}_{0}^{T} \otimes \mathbf{W}_{0}^{T}) \\
&= \lambda_{pr}^{2} (\mathbf{U}_{0} \otimes \mathbf{W}_{0}) \begin{bmatrix} \mathbf{I}_{r} \otimes [r \times r] & \mathbf{I}_{p-r} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{r} \\ \mathbf{I}_{p-r} \otimes \mathbf{I}_{r} \end{bmatrix} \\
&\times \begin{bmatrix} \mathbf{U}_{0}^{T} \otimes \mathbf{W}_{0}^{T} \\ \mathbf{U}_{0,\perp}^{T} \otimes \mathbf{W}_{0}^{T} \end{bmatrix} \\
&= \lambda_{pr}^{2} (\mathbf{U}_{0} \otimes \mathbf{W}_{0})(\mathbf{U}_{0}^{T} \otimes \mathbf{W}_{0}^{T}) \\
&+ \lambda_{pr}^{2} (\mathbf{U}_{0,\perp} \otimes \mathbf{W}_{0})(\mathbf{I}_{p-r} \otimes \mathbf{I}_{r}) (\mathbf{U}_{0,\perp}^{T} \otimes \mathbf{W}_{0}^{T}) \\
&= \lambda_{pr}^{2} (\mathbf{U}_{0} \mathbf{U}_{0}^{T} \otimes \mathbf{U}_{0,\perp} \mathbf{U}_{0,\perp}^{T} + \mathbf{U}_{0,\perp} \mathbf{U}_{0,\perp}^{T} \otimes \mathbf{U}_{0} \mathbf{U}_{0}^{T}) \\
&= \lambda_{pr}^{2} (\mathbf{U}_{0} \mathbf{U}_{0}^{T} \otimes (\mathbf{I}_{p} - \mathbf{U}_{0} \mathbf{U}_{0}^{T}) + (\mathbf{I}_{p} - \mathbf{U}_{0} \mathbf{U}_{0}^{T}) \otimes \mathbf{U}_{0} \mathbf{U}_{0}^{T}).
\end{align*}
\]
For the matrix $\Sigma_0^2 \otimes (I_p - U_0U_0^T) + (I_p - U_0U_0^T) \otimes \Sigma_0^2$, we write

$$\Sigma_0^2 \otimes (I_p - U_0U_0^T) + (I_p - U_0U_0^T) \otimes \Sigma_0^2$$

$$= W_0 \tilde{M}_0 W_0^T \otimes W_0 \begin{bmatrix} r & I_{p-r} \end{bmatrix} W_0^T + W_0 \begin{bmatrix} r & I_{p-r} \end{bmatrix} W_0^T \otimes W_0 \tilde{M}_0 W_0^T$$

$$= (W_0 \otimes W_0)(\tilde{V}_0 \otimes \tilde{V}_0) \left\{ \tilde{A}_0^2 \otimes \begin{bmatrix} r & I_{p-r} \end{bmatrix} + \begin{bmatrix} r & I_{p-r} \end{bmatrix} \otimes \tilde{A}_0^2 \right\}$$

$$\times (\tilde{V}_0 \otimes \tilde{V}_0)^T(W_0^T \otimes W_0^T)$$

$$= (W_0 \otimes W_0)(\tilde{V}_0 \otimes \tilde{V}_0) \left\{ \tilde{A}_0^2 \otimes \begin{bmatrix} r & I_{p-r} \end{bmatrix} \right\} \begin{bmatrix} I_{p-r} \otimes [\tilde{A}_0^2_{(p-r) \times (p-r)}] \\ I_{p-r} \otimes [\tilde{A}_0^2_{(p-r) \times (p-r)}] \end{bmatrix}$$

$$\times (\tilde{V}_0 \otimes \tilde{V}_0)^T(W_0^T \otimes W_0^T)$$

$$\geq \lambda_{0r}^2(W_0 \otimes W_0)(\tilde{V}_0 \otimes \tilde{V}_0) \left\{ I_r \otimes \begin{bmatrix} r & I_{p-r} \end{bmatrix} \right\} \begin{bmatrix} I_{p-r} \otimes [I_r_{(p-r) \times (p-r)}] \end{bmatrix}$$

$$= \lambda_{0r}^2 \left[ U_0 \otimes W_0 \ U_{0\perp} \otimes W_0 \right] \begin{bmatrix} r \times r \\ I_{p-r} \otimes [I_r_{(p-r) \times (p-r)}] \end{bmatrix} \left( W_0^T \otimes W_0^T \right)$$

$$\times \left[ U_0^T \otimes W_0^T \ U_{0\perp} \otimes W_0^T \right]$$

$$= \lambda_{0r}^2 \left( U_0 \otimes W_0 \right) \left\{ I_r \otimes \begin{bmatrix} I_{p-r} \end{bmatrix} \right\} \left( U_0^T \otimes W_0^T \right)$$

$$+ \lambda_{0r}^2 \left( U_{0\perp} \otimes W_0 \right) \left\{ I_{p-r} \otimes [I_r_{(p-r) \times (p-r)}] \right\} \left( U_{0\perp}^T \otimes W_0^T \right)$$

$$= \lambda_{0r}^2 \left( U_0 U_0^T \otimes U_{0\perp} U_{0\perp}^T \right) + \lambda_{0r}^2 \left( U_{0\perp} U_{0\perp}^T \otimes U_0 U_0^T \right)$$

and the proof is thus completed. □

**Proof of Lemma 3.8.** Let $U_{0\perp}$ be the orthogonal complement of $U_0$ such that $[U_0, U_{0\perp}] \in \mathbb{O}(p)$, and one can therefore take $W_0 = C_0^{-1}C_0$. For any $A \in \mathbb{R}^{(p-r) \times r}$ and $\varphi = vec(A)$,

$$vec(X_{\varphi})^T(C_0^T \otimes C_0^T)[U_0 U_0^T \otimes (I_p - U_0 U_0^T)](C_0 \otimes C_0)vec(X_{\varphi})$$

$$= \| (U_0^T \otimes U_{0\perp}^T)(C_0 \otimes C_0)vec(X_{\varphi}) \|_2^2$$

$$= \left\| \begin{bmatrix} C_{11} & C_{12}^T \ C_{21}^T \ C_{22}^T \end{bmatrix} \begin{bmatrix} r \times r \\ I_{p-r} \end{bmatrix} - A^T \right\|_2^2$$

$$= \left\| C_{22}^T A C_{11} - C_{12}^TA^TC_{21} \right\|_F^2,$$
and similarly,
\[
\text{vec}(X \varphi) \mathbin{\otimes} (C_0^T \otimes C_0^T) \{ (I_p - U_0^T U_0^T) \otimes U_0^T U_0^T \} (C_0 \otimes C_0) \text{vec}(X \varphi)
\]
\[
= \| (U_0^T U_0^T) (C_0 \otimes C_0) \text{vec}(X \varphi) \|^2_f
\]
\[
= \left\| \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} r \times r & -A^T \\ (p-r) \times (p-r) \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} r \times r \end{bmatrix} \right\|_F^2
\]
\[
= \| C_{22} A C_{11} - C_{12} A^T C_{21} \|^2_F,
\]
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.,
\[ p \geq 2r \text{ and } p < 2r. \]

**Case I:** \( p - r \geq r \). Let \( A_0 = U_{A_0} S_{A_0} V_{A_0}^T \) be the singular value decomposition of \( A_0 \), where \( U_{A_0} \in \mathbb{O}(p - r, r) \), \( V_{A_0} \in \mathbb{O}(r) \), and \( S_{A_0} = \text{diag}(\sigma_1(A_0), \ldots, \sigma_r(A_0)) \).

Suppose \( U_{A_0 \perp} \in \mathbb{O}(p - r, p - 2r) \) spans the orthogonal complement of \( \text{Span}(U_{A_0}) \), i.e., \( U_{A_0}^T U_{A_0 \perp} = r \times (p - 2r) \). Note that by assumption, \( 0 < \sigma_r(A_0) \leq \sigma_1(A_0) < 1 \).

Using the property of commutation matrices [43] that
\[
(C_{21} \otimes C_{12}) K_{(p-r)r} = K_{r(p-r)} (C_{12}^T \otimes C_{21}^T),
\]
we write
\[
(C_{21} \otimes C_{12}) K_{(p-r)r} = (V_{A_0} \otimes U_{A_0}) K_{rr},
\]
where
\[
\tilde{R} := (I_r + S_{A_0}^2)^{-1} \otimes \{ I_r - S_{A_0}(I_r + S_{A_0}^2)^{-1} S_{A_0} \}
\]
\[
+ K_{rr} \{ (I_r + S_{A_0}^2)^{-1} S_{A_0} \otimes S_{A_0}(I_r + S_{A_0}^2)^{-1} \}
\]
\[
= (I_r + S_{A_0}^2)^{-1} \otimes \{ I_r - S_{A_0}(I_r + S_{A_0}^2)^{-1} S_{A_0} \}
\]
\[
+ K_{rr} \{ (I_r + S_{A_0}^2)^{-1} S_{A_0} \otimes S_{A_0}(I_r + S_{A_0}^2)^{-1} \}.
\]
\[
\sigma_{\min} \{ (C_{21}^T \otimes C_{22}^T - (C_{21}^T \otimes C_{12}) K_{(p-r)r})
\]
\[
= \min \left[ \sigma_{\min}(\tilde{R}), \sigma_{\min}(\{ I_r + S_{A_0}^2 \}) \right].
\]
We now provide a lower bound for the smallest singular value of $\tilde{R}$. When $r = 1$, $K_{rr} = K_{11} = 1$, $S_{A_0} = \|A_0\|_2 \in \mathbb{R}^{1 \times 1}$, and hence

$$\tilde{R} = \frac{1}{(1 + S^2_{A_0})^2} + \frac{S^2_{A_0}}{(1 + S^2_{A_0})^2} = \frac{1}{1 + \|A\|_2^2}.$$  

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

$$D_1 = (I_r + S^2_{A_0})^{-1}, \quad D_2 = (I_r + S^2_{A_0})^{-1}S_{A_0}.$$

Note that by the property of the commutation matrix,

$$(K_{rr} + I_{r}) (D_2 \otimes D_2) = (D_2 \otimes D_2) (K_{rr} + I_{r}).$$

Therefore, $(K_{rr} + I_{r}) (D_2 \otimes D_2)$ is symmetric and is a product two positive semidefinite matrices $K_{rr} + I_{r}$ and $D_2 \otimes D_2$ because $K_{rr}$ is symmetric and only has eigenvalues in $\{0, 1\}$ [43]. Therefore, by Corollary 11 in [63], we have

$$\lambda_{\min}((K_{rr} + I_{r}) (D_2 \otimes D_2)) \geq \lambda_{\min}(K_{rr} + I_{r}) \lambda_{\min}(D_2 \otimes D_2) = 0.$$  

This further implies that $(K_{rr} + I_{r}) (D_2 \otimes D_2)$ is positive semidefinite and $\tilde{R}$ is symmetric. Note that $D_1 \succeq D_2$ because $\sigma_k(A_0) \in [0, 1]$, and that $\tilde{R}$ is positive semidefinite. Hence, to provide a lower bound for the smallest singular value of $\tilde{R}$, it is sufficient to provide a lower bound for the smallest eigenvalue of

$$(I_r + S^2_{A_0})^{-1} \otimes (I_r + S^2_{A_0})^{-1} - (I_r + S^2_{A_0})^{-1}S_{A_0} \otimes S_{A_0} (I_r + S^2_{A_0})^{-1}$$

$$= D_1 \otimes D_1 - D_2 \otimes D_2.$$  

Write

$$D_1 \otimes D_1 - D_2 \otimes D_2$$

$$= \text{diag} \left\{ \frac{1}{1 + \sigma^2_k(A_0)}, \ldots, \frac{1}{1 + \sigma^2_k(A_0)} \right\} \otimes \text{diag} \left\{ \frac{1}{1 + \sigma^2_r(A_0)}, \ldots, \frac{1}{1 + \sigma^2_r(A_0)} \right\}$$

$$- \text{diag} \left\{ \frac{\sigma_1(A_0)}{1 + \sigma^2_1(A_0)}, \ldots, \frac{\sigma_1(A_0)}{1 + \sigma^2_1(A_0)} \right\} \otimes \text{diag} \left\{ \frac{\sigma_r(A_0)}{1 + \sigma^2_r(A_0)}, \ldots, \frac{\sigma_r(A_0)}{1 + \sigma^2_r(A_0)} \right\}$$

$$= \begin{bmatrix} \frac{\sigma(A_0)}{1 + \sigma^2(A_0)} D_1 - D_2 & \cdots & D_1 - D_2 \end{bmatrix} \geq_{r^2 \times r^2}.$$

It follows that

$$\sigma_{\min}(\tilde{R}) = \sigma_{\min} \{ D_1 \otimes D_1 - D_2 \otimes D_2 + (K_{rr} + I_{r})(D_2 \otimes D_2) \}$$

$$= \lambda_{\min} \{ D_1 \otimes D_1 - D_2 \otimes D_2 + (K_{rr} + I_{r})(D_2 \otimes D_2) \}$$

$$\geq \lambda_{\min} \{ D_1 \otimes D_1 - D_2 \otimes D_2 \}$$

$$= \min_{k \in [r]} \lambda_{\min} \left\{ \frac{1}{1 + \sigma^2_k(A_0)} D_1 - \frac{\sigma_k(A_0)}{1 + \sigma^2_k(A_0)} D_2 \right\}$$

$$= \min_{k, l \in [r]} \left\{ \frac{1 - \sigma_k(A_0) \sigma_l(A_0)}{(1 + \sigma^2_k(A_0))(1 + \sigma^2_l(A_0))} \right\}. $$

This manuscript is for review purposes only.
We finally conclude that when \( r > 1 \),
\[
\sigma_{\min}\{C_{11} \otimes C_{22} - (C_{21}^T \otimes C_{12}^T)K_{(p-r)r}\} \\
= \lambda_{\min}\{C_{11} \otimes C_{22} - (C_{21}^T \otimes C_{12}^T)K_{(p-r)r}\} \\
= \min\left\{\sigma_{\min}(\bar{R}), \sigma_{\min}(D_1 \otimes I_{p-r})\right\} \\
\geq \min\left[\min_{k,l \in [r]} \frac{1 - \sigma_k(A_0)\sigma_l(A_0)}{(1 + \sigma_k^2(A_0))\{1 + \sigma_l^2(A_0)\}} \right], \\
\min_{k,l \in [r]} \frac{1}{1 + \sigma_k^2(A_0)} \\
= \frac{1}{1 + \|A_0\|_2^2},
\]
and when \( r = 1 \), we directly obtain
\[
\sigma_{\min}\{C_{11} \otimes C_{22} - (C_{21}^T \otimes C_{12}^T)K_{(p-r)r}\} \\
= \lambda_{\min}\{C_{11}^T \otimes C_{22}^T - (C_{21}^T \otimes C_{12}^T)^T K_{(p-r)r}\} \\
= \min\left\{\sigma_{\min}(\bar{R}), \sigma_{\min}(D_1 \otimes I_{p-r})\right\} \\
\geq \min\left[\min_{k,l \in [r]} \frac{1 - \sigma_k(A_0)\sigma_l(A_0)}{(1 + \sigma_k^2(A_0))\{1 + \sigma_l^2(A_0)\}} \right], \\
\min_{k,l \in [r]} \frac{1}{1 + \sigma_k^2(A_0)} \\
= \frac{1}{1 + \|A_0\|_2^2}.
\]

**Case II:** \( p - r < r \). This situation occurs only if \( r > 1 \). Let \( A_0 = U_{A_0}S_{A_0}V_{A_0}^T \) be the singular value decomposition of \( A_0 \), where \( U_{A_0} \in \mathbb{O}(p - r) \), \( V_{A_0} \in \mathbb{O}(r, p - r) \), and \( S_{A_0} = \text{diag}\{\sigma_1(A_0), \ldots, \sigma_{p-r}(A_0)\} \). Suppose \( V_{A_0\perp} \in \mathbb{O}(r, 2r - p) \) spans the orthogonal complement of \( \text{Span}(V_{A_0}) \), i.e., \( V_{A_0}^TV_{A_0\perp} = p-r\times(2r-p) \). Note that by assumption, \( 0 < \sigma_r(A_0) \leq \sigma_1(A_0) < 1 \). Observe that
\[
C_{11} = (I_r + A_{0}^T A_{0})^{-1} = [V_{A_0} \ V_{A_0\perp}] \left[\begin{array}{c}
(I_{p-r} + S_{A_0}^2)^{-1} \\
I_{2r-p} \\
\end{array}\right] [V_{A_0}^T \\
V_{A_0\perp}^T],
\]
and
\[
C_{22} = I_{p-r} - A_{0}(I_r + A_{0}^T A_{0})^{-1} A_{0}^T \\
= I_{p-r} - U_{A_0}S_{A_0}(I_{p-r} + S_{A_0}^2)^{-1} S_{A_0} U_{A_0}^T = U_{A_0}(I_{p-r} + S_{A_0}^2)^{-1} U_{A_0}^T.
\]

Also,
\[
C_{12} = -A_0(I_r + A_{0}^T A_{0})^{-1} A_{0}^T \\
= -U_{A_0}S_{A_0}(I_r + S_{A_0}^2)^{-1} V_{A_0}^T,
\]
\[
C_{21} = (I_r + A_{0}^T A_{0})^{-1} A_{0}^T = V_{A_0}(I_r + S_{A_0}^2)^{-1} S_{A_0} U_{A_0}^T.
\]

Using the property of commutation matrices [43] that
\[
(C_{21}^T \otimes C_{12}^T)K_{(p-r)r} = K_{r(p-r)}(C_{12}^T \otimes C_{21}^T),
\]
\[
K_{r(p-r)}(U_{A_0} \otimes V_{A_0}) = (V_{A_0} \otimes U_{A_0})K_{(p-r)(p-r)},
\]

This manuscript is for review purposes only.
we write
\[ C_{11} \otimes C_{22} - (C_{21}^T \otimes C_{12}^T)K_{(p-r)r} \]
\[ = C_{11} \otimes C_{22} - K_{(p-r)(C_{12}^T \otimes C_{21}^T)} \]
\[ = [V_{A_0} \otimes V_{A_0}] \left( [I_{p-r} + S_{A_0}^2]^{-1} \right) \left[ \begin{bmatrix} V_{A_0}^T \\ V_{A_0}^T \end{bmatrix} \otimes U_{A_0} \right] \left( [I_{p-r} + S_{A_0}^2]^{-1} \right) U_{A_0}^T \]
\[ + K_{(p-r)}(U_{A_0} \otimes S_{A_0}) \left[ [I_{p-r} + S_{A_0}^2]^{-1} V_{A_0}^T \otimes V_{A_0} \right] \left( [I_{p-r} + S_{A_0}^2]^{-1} S_{A_0} U_{A_0} \right) \]
\[ = \{V_{A_0}(I_{p-r} + S_{A_0})^{-1} V_{A_0}^T + V_{A_0} \otimes V_{A_0} \} \otimes \{U_{A_0}(I_{p-r} + S_{A_0})^{-1} S_{A_0} U_{A_0} \} \]
\[ + K_{(p-r)}(U_{A_0} \otimes V_{A_0}) \{S_{A_0}(I_{p-r} + S_{A_0})^{-1} \otimes (I_{p-r} + S_{A_0})^{-1} S_{A_0} \} \]
\[ \times (V_{A_0}^T \otimes U_{A_0}^T) \]
\[ = (V_{A_0} \otimes U_{A_0}) \{([I_{p-r} + S_{A_0}])^{-1} \otimes (I_{p-r} + S_{A_0})^{-1} \}(V_{A_0}^T \otimes U_{A_0}) \]
\[ + (V_{A_0} \otimes U_{A_0}) \{[I_{p-r} + S_{A_0}]^{-1} \otimes (I_{p-r} + S_{A_0})^{-1} \}(V_{A_0}^T \otimes U_{A_0}) \]
\[ + (V_{A_0} \otimes U_{A_0}) \{([I_{p-r} + S_{A_0}])^{-1} \otimes (I_{p-r} + S_{A_0})^{-1} \}(V_{A_0}^T \otimes U_{A_0}) \]
\[ = \tilde{R} = (I_{p-r} + S_{A_0})^{-1} \otimes (I_{p-r} + S_{A_0})^{-1} \]
\[ + K_{(p-r)(p-r)}([I_{p-r} + S_{A_0}]^{-1} S_{A_0} \otimes S_{A_0}) \]
\[ = (I_{p-r} + S_{A_0})^{-1} \otimes (I_{p-r} + S_{A_0})^{-1} \]
\[ + K_{(p-r)(p-r)}([I_{p-r} + S_{A_0}]^{-1} S_{A_0} \otimes S_{A_0}) \]
\[ = \sigma_{min}(C_{11} \otimes C_{22} - (C_{21}^T \otimes C_{12}^T)K_{(p-r)r}) = \sigma_{min}(\tilde{R}) \land \sigma_{min}(I_{2r-p} \otimes (I_{p-r} + S_{A_0})^{-1}) \]
\[ \text{where} \]
\[ \text{Note that} \{V_{A_0} \otimes U_{A_0}, V_{A_0} \otimes U_{A_0}\} \in \mathbb{O}(r(p-r)). \text{It follows that} \]
\[ \text{Similar to the case where} p - r \geq r, \text{we also provide a lower bound for the smallest}\]
\[ \text{Note that by the property of the commutation matrix,}\]
\[ (K_{(p-r)(p-r)} + I_{(p-r)^2}) \{([I_{p-r} + S_{A_0}]^{-1} S_{A_0} \otimes S_{A_0})(I_{p-r} + S_{A_0})^{-1} \}
\[ = (K_{(p-r)(p-r)} + I_{(p-r)^2})(D_2 \otimes D_2) = (D_2 \otimes D_2)(K_{(p-r)(p-r)} + I_{(p-r)^2}). \]
\[ \text{Therefore,} \{K_{(p-r)(p-r)} + I_{(p-r)^2})(D_2 \otimes D_2)\text{is symmetric and is a product two positive}\]
\[ \lambda_{min}\{K_{(p-r)(p-r)} + I_{(p-r)^2}\}(D_2 \otimes D_2) \]
\[ \geq \lambda_{min}(K_{(p-r)(p-r)} + I_{(p-r)^2}) \lambda_{min}(D_2 \otimes D_2) = 0. \]
\[ \tilde{R} \text{is symmetric. Note that} D_1 \succeq D_2 \text{because} \sigma_k(A_0) \in [0, 1], \text{and that} \tilde{R} \text{is positive}\]
semidefinite. Hence, to provide a lower bound for the smallest singular value of $\tilde{R}$, it is sufficient to provide a lower bound for the smallest eigenvalue of

$$
(I_{p-r} + S_{A_0}^2)^{-1} \otimes (I_{p-r} + S_{A_0}^2)^{-1} - (I_{p-r} + S_{A_0}^2)^{-1} S_{A_0} \otimes S_{A_0} (I_{p-r} + S_{A_0}^2)^{-1}
$$

Write

$$
D_1 \otimes D_1 - D_2 \otimes D_2
$$

$$
= \text{diag} \left\{ \frac{1}{1 + \sigma_1^2(A_0)}, \ldots, \frac{1}{1 + \sigma_1^2(A_0)} \right\}
$$

$$
\otimes \text{diag} \left\{ \frac{1}{1 + \sigma_1^2(A_0)}, \ldots, \frac{1}{1 + \sigma_1^2(A_0)} \right\}
$$

$$
- \text{diag} \left\{ \frac{\sigma_1(A_0)}{1 + \sigma_1^2(A_0)}, \ldots, \frac{\sigma_1(A_0)}{1 + \sigma_1^2(A_0)} \right\}
$$

$$
\otimes \text{diag} \left\{ \frac{\sigma_1(A_0)}{1 + \sigma_1^2(A_0)}, \ldots, \frac{\sigma_1(A_0)}{1 + \sigma_1^2(A_0)} \right\}
$$

$$
= \begin{bmatrix}
1 - \frac{\sigma_1(A_0)}{1 + \sigma_1^2(A_0)} & D_1 - \frac{\sigma_1(A_0)}{1 + \sigma_1^2(A_0)} & D_2 \\
- \frac{1}{1 + \sigma_1^2(A_0)} & \ddots & - \frac{1}{1 + \sigma_1^2(A_0)} \\
- \frac{1}{1 + \sigma_1^2(A_0)} & - \frac{1}{1 + \sigma_1^2(A_0)} & - \frac{1}{1 + \sigma_1^2(A_0)} \\
\end{bmatrix}
$$

$$
\geq p-r^2 \times p-r^2.
$$

It follows that

$$
\sigma_{\min}(\tilde{R}) = \sigma_{\min} \left\{ D_1 \otimes D_1 - D_2 \otimes D_2 + (K_{(p-r)(p-r)} + I_{(p-r)^2})(D_2 \otimes D_2) \right\}
$$

$$
= \lambda_{\min} \left\{ D_1 \otimes D_1 - D_2 \otimes D_2 + (K_{(p-r)(p-r)} + I_{(p-r)^2})(D_2 \otimes D_2) \right\}
$$

$$
\geq \lambda_{\min} \left\{ D_1 \otimes D_1 - D_2 \otimes D_2 \right\}
$$

$$
= \min_{k \in [p-r]} \lambda_{\min} \left\{ \frac{1}{1 + \sigma_k^2(A_0)} D_1 - \frac{\sigma_k(A_0)}{1 + \sigma_k^2(A_0)} D_2 \right\}
$$

$$
= \min_{k,l \in [p-r]} \frac{1 - \sigma_k(A_0) \sigma_l(A_0)}{\left\{ 1 + \sigma_k^2(A_0) \right\} \left\{ 1 + \sigma_l^2(A_0) \right\}}.
$$

We finally conclude that

$$
\sigma_{\min} \left\{ C_{11} \otimes C_{22} - (C_{21}^T \otimes C_{12}^T) K_{(p-r)r} \right\}
$$

$$
= \lambda_{\min} \left\{ C_{11} \otimes C_{22} - (C_{21}^T \otimes C_{12}^T) K_{(p-r)r} \right\}
$$

$$
= \min \left\{ \sigma_{\min}(\tilde{R}), \sigma_{\min}(I_{2r-r} \otimes D_1) \right\}
$$

$$
\geq \min \left[ \min_{k,l \in [p-r]} \frac{1 - \sigma_k(A_0) \sigma_l(A_0)}{\left\{ 1 + \sigma_k^2(A_0) \right\} \left\{ 1 + \sigma_l^2(A_0) \right\}}, \frac{1}{\left\{ 1 + \|A_0\|_2^2 \right\}^2} \right].
$$

The proof is thus completed. □

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:

This manuscript is for review purposes only.
For every $\Sigma \in \mathcal{M}(p, r)$ and $\Pi \in \mathcal{G}(p, r)$, there exist neighborhoods $\mathcal{V}_1$ of $\Sigma$ and $\mathcal{V}_2$ of $\Pi$ in $\mathbb{R}^{p \times p}$ and permutation matrices $\mathbf{P}_1, \mathbf{P}_2 \in \mathcal{P}(p)$, such that the maps $x_{\mathbf{P}_1} : \mathcal{U} := \mathcal{D}(p, r) \rightarrow \mathcal{M}(p, r) \cap \mathcal{V}$ onto $\mathcal{M}(p, r) \cap \mathcal{V}_1$ and $y_{\mathbf{P}_2} : \mathcal{A} \rightarrow \mathcal{G}(p, r) \cap \mathcal{V}_2$ onto $\mathcal{G}(p, r) \cap \mathcal{V}_2$ satisfy

(a) $x_{\mathbf{P}_1}$ and $y_{\mathbf{P}_2}$ are differentiable and are homeomorphisms;

(b) The Jacobians $\partial \text{vec}(x_{\mathbf{P}_1}(\theta)) / \partial \theta^T$ and $\partial \text{vec}(y_{\mathbf{P}_2}(\varphi)) / \partial \varphi^T$ have ranks $d$ and $(p - r)$, 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 $\mathbf{S}_\Sigma$ is the diagonal matrix of the nonzero eigenvalues of $\Sigma$ and $\mathbf{U}_\Sigma \in \mathcal{O}(p, r)$. Similarly, let $\Pi_0 \in \mathcal{G}(p, r)$ with $\Pi_0 = \mathbf{Q}_\Pi \mathbf{Q}_\Pi^T$ for some $\mathbf{Q} \in \mathcal{O}(p, r)$. Clearly, there exist permutation matrices $\mathbf{P}_1, \mathbf{P}_2 \in \{0, 1\}^{p \times p}$ and $\mathbf{V}_1, \mathbf{V}_2 \in \mathcal{O}(r)$ such that $\mathbf{P}_1^T \mathbf{U}_\Sigma \mathbf{V}_1, \mathbf{P}_2^T \mathbf{Q} \mathbf{V}_2 \in \mathcal{O}_+(p, r)$. This implies that $\mathbf{P}^T \Sigma_0 \mathbf{P} \in \mathcal{S}(p, r)$ and hence, there exists $\mathbf{A}_0, \mathbf{A}_2 \in \mathbb{R}^{(p - r) \times r}$ with $\|\mathbf{A}_0\|_2, \|\mathbf{A}_2\|_2 < 1$ and $\theta_0 = \text{vec}(\mathbf{A}_0)$, $\varphi_2 = \text{vec}(\mathbf{A}_2) \in \mathcal{A}$, $\theta_0 = [\theta_0^T, \varphi_2^T] \in \mathcal{D}(p, r)$, such that $\Sigma_0 = \mathbf{P}_1 \Sigma(\theta_0) \mathbf{P}_1^T$ and $\Pi_0 = \mathbf{P}_2 \mathbf{U}(\varphi_2) \mathbf{U}(\varphi_2)^T \mathbf{P}_2^T$. We next verify that $x_{\mathbf{P}_1}$ and $y_{\mathbf{P}_2}$ have the desired properties.

(i) Clearly, $x_{\mathbf{P}_1}$ is differentiable since $\text{vec}(x_{\mathbf{P}_1}(\theta)) = (\mathbf{P}_1^T \otimes \mathbf{P}_1^T) \text{vec}\{\Sigma(\theta)\}$ and $\Sigma(\cdot)$ is differentiable. Similarly, $y_{\mathbf{P}_2}$ is also differentiable.

(ii) We show that the image of $x_{\mathbf{P}_1} : \mathcal{U} \rightarrow \mathcal{M}(p, r)$ 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 $y_{\mathbf{P}_2} : \mathcal{A} \rightarrow \mathcal{G}(p, r)$ is given by

$$\{\mathbf{P}_2 \mathbf{U} \mathbf{U}^T \mathbf{P}_2^T : \mathbf{U} \in \mathcal{O}_+(p, r)\}.$$
\[ \theta = [\varphi_1^T, \mu^T]^T, \text{ where } \varphi_1 = \text{vec}(A_1) \text{ for some } \|A\|_2 < 1, A \in \mathbb{R}^{(p-r) \times r}. \]

Let

\[ \delta_1 = \min \left\{ \frac{\|\Sigma(\theta)\|_2, \min_{k \in [r]} \lambda_k^2(\Sigma(\theta)) \left(1 - \|A_1\|_2^2\right)}{12\sqrt{2} \|\Sigma(\theta)\|_2 (1 + \|A_1\|_2^2)} \right\} > 0, \]

\[ \delta_2 = \frac{(1 - \|A_1\|_2^2)}{8\sqrt{2} (1 + \|A_1\|_2^2)} > 0, \]

and consider any \( \Sigma' \in \mathbb{P}_{\mathcal{Y}(p,r)} \) with \( \|\Sigma' - P\Sigma(\theta)P^T\|_F < \delta_1 \) and \( QQ^T \in \mathcal{G}(p,r) \) with \( \|QQ^T - P_2U(\varphi_2)U(\varphi_2)^TP_2^T\|_F < \delta_2 \), where \( Q \in \mathbb{O}(p,r) \). It is sufficient to show that \( P_1^T\Sigma P_1 \in \mathcal{G}(p,r) \) and \( QQ^T \in \mathcal{Y}_p(A) \).

For the first part, let \( U'S'(U')^T \) be the eigendecomposition of \( P_1^T\Sigma P_1 \), where \( U' \in \mathbb{O}(p,r) \) and \( S' \) is the diagonal matrix of the nonzero eigenvalues of \( P_1^T\Sigma P_1 \). By Theorem 3 in [62], there exists \( O \in \mathbb{O}(r) \) such that

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

Let \( Q_1 \) be the top square block of \( U' \), i.e., \( Q_1 = I_{p \times r} \). It follows from Weyl's inequality for singular values that

\[ |\sigma_r(Q_1) - \sigma_r(I_{p \times r}^T U(\varphi))| \leq \|Q_1O - I_{p \times r}^T U(\varphi)\|_2 \leq \|U'O - U(\varphi)\|_F \]

\[ \leq \frac{6\sqrt{2} \|\Sigma(\theta)\|_2 \delta}{\min_{k \in [r]} \lambda_k^2(\Sigma(\theta))} \leq \frac{1 - \|A_1\|_2^2}{2(1 + \|A_1\|_2^2)}, \]

Since \( \sigma_r(I_{p \times r}^T U(\varphi)) = \sigma_r((I_r - A^TA)(I_r + A^TA)^{-1}) = (1 - \|A_1\|_2^2)/(1 + \|A_1\|_2^2) \), we obtain that

\[ \sigma_r(Q_1) \geq \frac{1 - \|A_1\|_2^2}{2(1 + \|A_1\|_2^2)} > 0. \]

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

\[ P_1^T\Sigma P_1 = U'S'(U')^T = (U'\mathcal{V}')((U')^T\mathcal{V}'(U')^T) \in \mathcal{G}(p,r). \]

This completes the proof that \( P_1 \in \mathcal{G}(p,r) \) is open with regard to the Euclidean (Frobenius) topology restricted to \( \mathcal{G}(p,r) \).

To show that \( QQ^T \in \mathcal{Y}_p(A) \), we apply Davis-Kahan theorem (see Theorem 2 in [62]) and obtain

\[ \|P_2^TQO' - U(\varphi)\|_F \leq 2\sqrt{2} \|QQ^T - P_2U(\varphi_2)U(\varphi_2)^TP_2^T\|_F \]

\[ \leq 2\sqrt{2} \delta_2 < \frac{1 - \|A_2\|_2^2}{4(1 + \|A_2\|_2^2)} = \frac{1}{4} \sigma_r(I_{p \times r}^T U(\varphi)) \]

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

\[ \sigma_r(I_{p \times r}^T P_2^T Q) > \frac{1}{2} \sigma_r(I_{p \times r}^T Q) > 0, \]

and hence, there exists some \( \mathcal{V}'' \in \mathbb{O}(r) \) such that \( P_2^TQ\mathcal{V}'' \in \mathbb{O}_+(p,r) \). Therefore,

\[ QQ^T = P_2(P_2^TQ\mathcal{V}'')(P_2^TQ\mathcal{V}'')^TP_2^T \in \mathcal{Y}_p(A). \]

This shows that \( \mathcal{Y}_p(A) \) is open with respect to the Euclidean topology restricted to \( \mathcal{G}(p,r) \).
We show that $x_P^1 : \mathcal{U} \to x_P^1(\mathcal{U})$ and $y_P^2 : \mathcal{A} \to y_P^2(\mathcal{A})$ are homeomorphisms. This is equivalent to show that $x_P^1$ and $y_P^2$ are continuous and they have continuous inverses $x_P^{-1} : x_P^1(\mathcal{U}) \to \mathcal{U}$ and $y_P^{-1} : y_P^2(\mathcal{A}) \to \mathcal{A}$, respectively. Now suppose $\theta = [\theta_1, \ldots, \theta_d]^T$ and $\varphi = [\phi_1, \ldots, \phi_s]$, $s = (p - r)r$, and write

$$\text{vec}\{x_P^1(\theta)\} = [\sigma_1(\theta_1, \ldots, \theta_d), \ldots, \sigma_p^2(\theta_1, \ldots, \theta_d)]^T,$$

$$\text{vec}\{y_P^2(\varphi)\} = [\pi_1(\phi_1, \ldots, \phi_s), \ldots, \pi_p^2(\phi_1, \ldots, \phi_s)]^T.$$

By Theorem 3.6 and the fact that

$$\text{vec}\{x_P(\theta)\} = (P \otimes P)\text{vec}\{\Sigma(\theta)\},$$

$$\text{vec}\{y_P(\varphi)\} = (P \otimes P)\text{vec}\{U(\varphi)U(\varphi)^T\},$$

we may assume the Jacobian matrices

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

are non-singular without loss of generality. Consider extensions of $\text{vec}\{x_P^1(\cdot)\}$ and $\text{vec}\{y_P^2(\cdot)\}$ defined by

$$g : \mathcal{U}(p, r) \times \mathbb{R}^{p^2 - d} \to \mathbb{R}^{p^2},$$

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

where

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

and

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

Then the Jacobian matrix of $g$ with respect to the vector

$$[\theta_1, \ldots, \theta_d, \theta_{d+1}, \ldots, \theta_{p^2}]^T$$
is also non-singular at $[\theta^T, \theta_{d+1}, \ldots, \theta_{p^2}]^T$ for any $[\theta_{d+1}, \ldots, \theta_{p^2}]^T \in \mathbb{R}^{p^2-d}$.

Similarly, the Jacobian matrix of $h$ with respect to the vector $[\phi_1, \ldots, \phi_s, \phi_{s+1}, \ldots, \phi_{s^2}]^T$

is also non-singular at $[\varphi^T, \phi_{s+1}, \ldots, \phi_{s^2}]^T$ for any $[\phi_{s+1}, \ldots, \phi_{s^2}]^T \in \mathbb{R}^{p^2-s}$.

By the inverse mapping theorem, the function

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

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

has a continuous inverse

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

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

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

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

has a continuous inverse

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

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

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

Then we claim that $\pi_d \circ g^{-1}|_{\text{vec}(\xi_{p^2})} : U \rightarrow \mathbb{R}^{p^2}$ is the inverse of $\text{vec}(\xi_{p^2}(\cdot))$, and $\pi_s \circ h^{-1}|_{\text{vec}(\eta_{p^2})} : A \rightarrow \mathbb{R}^{p^2}$ is the inverse of $\text{vec}(\eta_{p^2}(\cdot))$. In fact, for any $\theta = [\theta_1, \ldots, \theta_d]^T \in \mathcal{D}(p, r)$ and $\varphi = [\phi_1, \ldots, \phi_s]^T \in A$, by the facts that

$$\text{vec}(\xi_{p^2}(\theta)) = g(\theta_1, \ldots, \theta_d, 0, \ldots, 0),$$

$$\text{vec}(\eta_{p^2}(\varphi)) = h(\phi_1, \ldots, \phi_s, 0, \ldots, 0),$$

we have

$$(\pi_d \circ g^{-1}|_{\text{vec}(\xi_{p^2})}(\varphi)) \circ \xi_{p^2}(\theta) = \{\pi_d \circ g^{-1} \circ \text{vec}(\xi_{p^2})\}(\theta)$$

$$= (\pi_d \circ g^{-1} \circ g)(\theta_1, \ldots, \theta_d, 0, \ldots, 0)$$

$$= \pi_d(\theta_1, \ldots, \theta_d, 0, \ldots, 0) = \theta,$$

$$(\pi_s \circ h^{-1}|_{\text{vec}(\eta_{p^2})}(\varphi)) \circ \eta_{p^2}(\theta) = \{\pi_s \circ h^{-1} \circ \text{vec}(\eta_{p^2})\}(\varphi)$$

$$= (\pi_s \circ h^{-1} \circ h)(\phi_1, \ldots, \phi_s, 0, \ldots, 0)$$

$$= \pi_s(\phi_1, \ldots, \phi_d, 0, \ldots, 0) = \theta.$$

Also, because $\pi_d$, $\pi_s$, $g^{-1}$, and $h^{-1}$ are continuous, it follows that the composite restrictions $\pi_d \circ g^{-1}|_{\text{vec}(\xi_{p^2})}(U)$ and $\pi_s \circ h^{-1}|_{\text{vec}(\eta_{p^2})}(A)$ are also continuous. This shows that $\pi_d \circ g^{-1}|_{\text{vec}(\xi_{p^2})}(U)$ and $\pi_s \circ h^{-1}|_{\text{vec}(\eta_{p^2})}(A)$ are the continuous inverses of $\text{vec}(\xi_{p^2}(\cdot))$ and $\text{vec}(\eta_{p^2}(\cdot))$. It is also clear that $\xi_{p^2}$ and $\eta_{p^2}$ are continuous. Hence, we conclude that $\xi_{p^2} : U \rightarrow \xi_{p^2}(U)$ and $\eta_{p^2} : A \rightarrow \eta_{p^2}(A)$ are homeomorphisms.
(v) We show that Jacobians $\partial \text{vec}\{x_P, (\theta)\}/\partial \theta^T$ and $\partial \text{vec}\{y_P, (\varphi)\}/\partial \varphi^T$ have ranks $d$ and $s$, respectively. This is immediate by Theorem 3.6 and the fact that

$$\text{vec}\{x_P, (\theta)\} = (P_1 \otimes P_1) \text{vec}\{\Sigma(\theta)\},$$

$$\text{vec}\{y_P, (\varphi)\} = (P_2 \otimes P_2) \text{vec}(U(\varphi)U(\varphi)^T).$$

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

Derivation of Example 4.3. To verify that $\gamma(\cdot)$ given above is indeed a geodesic, we consider a local coordinate system $(A, y_P)$ with $y_P(\varphi) = uu^T, \varphi \in \mathbb{R}^{p-1}, \|\varphi\|_2 < 1$, and $P$ is a permutation matrix such that $P^Tu = O(p, 1)$, where $v \in \{\pm 1\}$ adjusts the sign of the first coordinate of $P^Tu$. Without loss of generality, we may assume that $v = +1$ because $v$ does not affect the formula of $\gamma(\cdot)$. Since $\Xi \in T_1G(p, 1)$, it can be written as $\Xi = U_{\perp}zu^T + zu^T U_{\perp}^T$ for some $z \in \mathbb{R}^{p-1}$, where $U_{\perp} \in O(p, p-1)$ spans the orthogonal complement of $\text{Span}(u)$. Let $\sigma := \|\Xi u\|_2, u = \Xi u/\sigma$ and write $P^Tu, P^T\dot{u}$ in the following block form

$$P^Tu = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad P^T\dot{u} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix},$$

where $u_1, \delta_1 \in \mathbb{R}, u_1 > 0$, and $u_2, \delta_2 \in \mathbb{R}^{p-1}$. Consider a curve $\varphi(\cdot) : (-\epsilon, \epsilon) \to A = \{\varphi \in \mathbb{R}^{p-1} : \|\varphi\|_2 < 1\}$ given by

$$\varphi(t) = \frac{u_2 \cos \sigma t + \delta_2 \sin \sigma t}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t}.$$

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

$$\frac{1 - \varphi(t)^T \varphi(t)}{1 + \varphi(t)^T \varphi(t)} = u_1 \cos \sigma t + \delta_1 \sin \sigma t, \quad \frac{2\varphi(t)}{1 + \varphi(t)^T \varphi(t)} = u_2 \cos \sigma t + \delta_2 \sin \sigma t.$$

It follows that

$$y_P(\varphi(t)) = P \begin{bmatrix} u_1 \cos \sigma t + \delta_1 \sin \sigma t \\ u_2 \cos \sigma t + \delta_2 \sin \sigma t \end{bmatrix} \begin{bmatrix} u_1 \cos \sigma t + \delta_1 \sin \sigma t \\ u_2^2 \cos \sigma t + \delta_2^2 \sin \sigma t \end{bmatrix} P^T$$

$$= (u \cos \sigma t + \dot{u} \sin \sigma t)(u \cos \sigma t + \dot{u} \sin \sigma t)^T = \gamma(t).$$

It is now sufficient to verify that $\varphi(t)$ satisfies the geodesic equation under the local coordinate $(A, y_P)$. By Propositions 4.2, the Gram matrix of the Riemannian metric at $\Pi = y_P(\varphi)$ under $(A, y_P)$ is

$$[g_{ij}]_{(p-1) \times (p-1)} = \frac{1}{1 + \varphi^T \varphi} \left( I_{p-1} - \frac{\varphi \varphi^T}{1 + \varphi^T \varphi} \right) = \frac{I_{p-1}}{(1 + \varphi^T \varphi)^2}.$$

Let $[g'_{ij}]_{(p-1) \times (p-1)}$ be the inverse of $[g_{ij}]_{(p-1) \times (p-1)}$ and write $\varphi = [\phi_1, \ldots, \phi_{p-1}]^T$. The Christoffel symbol $\Gamma^m_{ij}$ for any $i, j, m \in \{p - 1\}$ can be computed as

$$\Gamma^m_{ij} = \frac{1}{2} \sum_{k=1}^{p-1} \left( \frac{\partial g_{jk}}{\partial \phi_i} + \frac{\partial g_{ik}}{\partial \phi_j} - \frac{\partial g_{ij}}{\partial \phi_k} \right) g^{km}$$

$$= \mathbb{I}(i = j) \frac{2\phi_m}{1 + \varphi^T \varphi} - \mathbb{I}(i = m) \frac{2\phi_j}{1 + \varphi^T \varphi} - \mathbb{I}(j = m) \frac{2\phi_i}{1 + \varphi^T \varphi}.$$
Write $\varphi(t) = [\varphi_1(t), \ldots, \varphi_{p-1}(t)]^T$. We now argue that for each $m \in [p-1]$, $\varphi_m(t)$ satisfies the geodesic equation

$$\varphi_m''(t) + \sum_{i=1}^{p-1} \sum_{j=1}^{p-1} \Gamma_{ij}^m \varphi_i'(t) \varphi_j'(t) = 0.$$  

Write $u_2 = [u_{21}, \ldots, u_{2(p-1)}]^T$ and $\delta_2 = [\delta_{21}, \ldots, \delta_{2(p-1)}]^T$. A direct computation yields

$$\varphi'_m(t) = \frac{-\sigma u_{2m} \sin \sigma t + \sigma \delta_{2m} \cos \sigma t}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} - \frac{(u_{2m} \cos \sigma t + \delta_{2m} \sin \sigma t)(-\sigma u_1 \sin \sigma t + \sigma_1)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2},$$

and

$$\frac{2 \varphi_i(t) \varphi_i'(t)}{1 + \varphi(t)^T \varphi(t)} = \frac{(-\sigma u_{21}^2 + \sigma \delta_{21}^2) \sin \sigma t \cos \sigma t - \sigma u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t)}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} - \frac{(u_{21}^2 \cos^2 \sigma t + \delta_{21}^2 \sin^2 \sigma t - 2 u_1 \delta_1 \sin \sigma t \cos \sigma t)(-\sigma u_1 \sin \sigma t + \sigma_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2}.$$

It follows that

$$\sum_{i=1}^{p-1} \frac{2 \varphi_i(t) \varphi_i'(t)}{1 + \varphi(t)^T \varphi(t)} = \frac{(\delta_1 - \delta_2) \sin \sigma t \cos \sigma t - \sigma u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t)}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} - \frac{(\delta_2 - \delta_1) \sin \sigma t \cos \sigma t - \sigma u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t)}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} + \sigma u_1 \sin \sigma t - \sigma_1 \cos \sigma t \frac{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t}. $$

Therefore,

$$2 \varphi''_m(t) \sum_{i=1}^{p-1} \frac{2 \varphi_i(t) \varphi_i'(t)}{1 + \varphi(t)^T \varphi(t)} = -\frac{2(\sigma u_{2m} \sin \sigma t - \sigma \delta_{2m} \cos \sigma t)(\sigma u_1 \sin \sigma t - \sigma_1)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} + \frac{2(u_{2m} \cos \sigma t + \delta_{2m} \sin \sigma t)(\sigma u_1 \sin \sigma t - \sigma_1 \cos \sigma t)^2}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^3}. $$
Also, another direct computation shows that
\[
\sum_{i=1}^{p-1} \frac{2\{\phi'_i(t)\}^2}{1 + \varphi(t)^2 \varphi(t)} = \frac{\sigma^2(1 - u_1^2) \sin^2 \sigma t + \sigma^2(1 - \delta_1^2) \cos^2 \sigma t + 2\sigma^2 u_1 \delta_1 \sin \sigma t \cos \sigma t}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} 
\]
\[
+ \frac{\sigma^2 u_1^2 \sin^2 \sigma t + \sigma^2 \delta_1^2 \cos^2 \sigma t - 2\sigma^2 u_1 \delta_1 \sin \sigma t \cos \sigma t}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^3} \{1 - (u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2\} 
\]
\[
+ \frac{2(\sigma u_1 \sin \sigma t - \sigma \delta_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} \{\sigma(u_1^2 - \delta_1^2) \sin \sigma t \cos \sigma t - \sigma u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t)\} 
\]
\[
= \frac{\sigma^2 \{1 - (u_1 \sin \sigma t - \delta_1 \cos \sigma t)^2\}}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} 
\]
\[
+ \frac{\sigma^2 (u_1 \sin \sigma t - \delta_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} \times \{u_1 \sin \sigma t - \delta_1 \cos \sigma t + (\delta_1^2 - u_1^2) \sin \sigma t \cos \sigma t + u_1 \delta_1 (\sin^2 \sigma t - \cos^2 \sigma t)\} 
\]
\[
+ \frac{\sigma^2 (u_1 \sin \sigma t - \delta_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} \times \{2(u_1^2 - \delta_1^2) \sin \sigma t \cos \sigma t - 2\sigma u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t)\} 
\]
\[
= \frac{\sigma^2 \{1 - (u_1 \sin \sigma t - \delta_1 \cos \sigma t)^2\}}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} \times \{u_1 \sin \sigma t - \delta_1 \cos \sigma t + (u_1^2 - \delta_1^2) \sin \sigma t \cos \sigma t + u_1 \delta_1 (\cos^2 \sigma t - \sin^2 \sigma t)\} 
\]
\[
+ \frac{\sigma^2 (u_1 \sin \sigma t - \delta_1 \cos \sigma t)^2}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} + \frac{\sigma^2 (u_1 \sin \sigma t - \delta_1 \cos \sigma t)^2}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t} 
\]
\[
= \frac{\sigma^2}{1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t}. 
\]

Hence,
\[
(A.2) \quad \phi_m(t) = \sum_{i=1}^{p-1} \frac{2\{\phi'_i(t)\}^2}{1 + \varphi(t)^2 \varphi(t)} \frac{\sigma^2 (u_{2m} \cos \sigma t + \delta_{2m} \sin \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2}.
\]

Using the formula for the Christoffel symbols, the geodesic equation becomes
\[
\phi''_m(t) = 2\phi'_m(t) \sum_{i=1}^{p-1} \frac{2\phi_i(t) \phi'_i(t)}{1 + \varphi(t)^2 \varphi(t)} - \sum_{i=1}^{p-1} \frac{2\phi'_m(t) \phi'_i(t)}{1 + \varphi(t)^2 \varphi(t)}. 
\]

This manuscript is for review purposes only.
A direct computation shows that
\[
\phi_m''(t) = \frac{-\sigma^2(u_{2m} \cos \sigma t + \delta_{2m} \sin \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} 
- \frac{2(-\sigma u_{2m} \sin \sigma t + \sigma \delta_{2m} \cos \sigma t)(-\sigma u_1 \sin \sigma t + \sigma \delta_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^2} 
+ \frac{2(u_{2m} \cos \sigma t + \delta_{2m} \sin \sigma t)(-\sigma u_1 \sin \sigma t + \sigma \delta_1 \cos \sigma t)}{(1 + u_1 \cos \sigma t + \delta_1 \sin \sigma t)^3}.
\]

Hence, the verification that \( \varphi(t) \) satisfies the above equation is completed by combining the results in (A.1), (A.2), and the formula for \( \phi_m''(t) \) in (A.3).

**Proof of Proposition 4.4.** Since \( \Sigma(\theta) = D \Sigma(\theta_0)D^T \), it follows that the leading eigenvector \( u \) of \( \Sigma(\theta) \) has the form \( u = Du_0v/\|Du_0\|_2 \), where \( v \in \{ \pm 1 \} \) adjusts for the sign of the first element of \( Du_0 \) such that \( Du_0/\|Du_0\|_2 \in \mathbb{O}_+(p, 1) \). Write \( D = I_p \) in the block form
\[
D - I_p = \begin{bmatrix} \epsilon & \alpha^T \\ \beta & \Gamma \end{bmatrix},
\]
where \( \epsilon \in \mathbb{R}, \alpha, \beta \in \mathbb{R}^{p-1} \), and \( \Gamma \in \mathbb{R}^{(p-1) \times (p-1)} \). Because the first element of \( Du_0 \) can be written as
\[
[1 + \epsilon \alpha^T]_0 = \frac{(1 + \epsilon)(1 - \varphi_0^T \varphi_0) + 2\alpha^T \varphi_0}{1 + \varphi_0^T \varphi_0} \geq 1 - \varphi_0^T \varphi_0 - |\epsilon| - \|\alpha\|_2
\]
\[
\geq 1 - \varphi_0^T \varphi_0 - 2|\|D - I_p\|_F| \geq \frac{1 - \varphi_0^T \varphi_0}{2(1 + \varphi_0^T \varphi_0)} > 0.
\]
This implies that \( v = +1 \) and we can remove \( v \) from the expression of \( u \). Also, observe that
\[
Du_0 = \begin{bmatrix} (1 + \epsilon) \left( \frac{1 - \varphi_0^T \varphi_0}{1 + \varphi_0^T \varphi_0} \right) + 2\alpha^T \varphi_0 \\ \beta \left( \frac{1 - \varphi_0^T \varphi_0}{1 + \varphi_0^T \varphi_0} \right) + 2(I_{p-1} + \Gamma) \varphi_0 \\ (1 + \varphi_0^T \varphi_0) \|Du_0\|_2 + (1 + \epsilon)(1 - \varphi_0^T \varphi_0) + 2\alpha^T \varphi_0 \end{bmatrix}.
\]
By the inverse Cayley parameterization (2.4), we have
\[
\varphi = \frac{\beta(1 - \varphi_0^T \varphi_0) + 2(I_{p-1} + \Gamma) \varphi_0}{\|Du_0\|_2(1 + \varphi_0^T \varphi_0)} \left( 1 + \frac{(1 - \varphi_0^T \varphi_0)(1 + \epsilon) + 2\alpha^T \varphi_0}{\|Du_0\|_2(1 + \varphi_0^T \varphi_0)} \right)^{-1}
\]
\[
= \frac{\beta(1 - \varphi_0^T \varphi_0) + 2(I_{p-1} + \Gamma) \varphi_0}{(1 + \varphi_0^T \varphi_0) \|Du_0\|_2 + (1 + \epsilon)(1 - \varphi_0^T \varphi_0) + 2\alpha^T \varphi_0}.
\]
Note that the denominator satisfies
\[
(1 + \varphi_0^T \varphi_0) \|Du_0\|_2 + (1 + \epsilon)(1 - \varphi_0^T \varphi_0) + 2\alpha^T \varphi_0
\]
\[
= 2 + (1 + \varphi_0^T \varphi_0)(\|Du_0\|_2 - 1) + \epsilon(1 - \varphi_0^T \varphi_0) + 2\alpha^T \varphi_0.
\]
It follows that
\[
(1 + \varphi_0^T \varphi_0) \|Du_0\|_2 + (1 + \epsilon)(1 - \varphi_0^T \varphi_0) + 2\alpha^T \varphi_0 \geq 2 - 4\|D - I_p\|_F
\]
and
\[
\|\varphi - \varphi_0\|_2 \leq \frac{(1 + \varphi_0^T \varphi_0) \|Du_0\|_2 - 1}{2 - 4\|D - I_p\|_F} + \frac{(1 + \varphi_0^T \varphi_0)(|\epsilon| + \|\beta\|_2)}{2 - 4\|D - I_p\|_F} + \frac{2\|I\varphi_0 - \varphi_0 \alpha^T \varphi_0\|_2}{2 - 4\|D - I_p\|_F}.
\]

\[\square\]
REFERENCES


46  F. XIE


