Gaussian and Robust Kronecker Product Covariance Estimation: Existence and Uniqueness

Ilya Soloveychik and Dmitry Trushin, Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel

Abstract

We consider Gaussian and robust covariance estimation assuming the true covariance matrix to be the Kronecker product of two lower dimensional square matrices. In both cases we define the estimators as solutions to constrained maximum likelihood programs. In the robust case we consider Tyler’s estimator defined as a maximum likelihood estimator of a certain distribution on a sphere. We develop tight sufficient conditions for the existence and uniqueness of the estimates and show that in Gaussian case with the unknown mean \( \frac{p}{q} + \frac{q}{p} + 2 \) is almost surely enough to guaranty the existence and uniqueness, where \( p \) and \( q \) are the dimensions of the Kronecker product factors of the true covariance. In robust case with the mean known the corresponding sufficient number of samples is \( \max \left[ \frac{p}{q}, \frac{q}{p} \right] + 1 \).

Index Terms

Constrained covariance estimation, robust covariance estimation, high-dimensional covariance estimation, Kronecker product structure.

I. INTRODUCTION

Covariance estimation is a fundamental problem in multivariate statistical analysis. It has received attention in diverse fields including economics and financial time series analysis (e.g., portfolio selection, risk management and asset pricing [1]), bioinformatics (e.g., gene microarray data [2, 3], functional MRI [4]) and machine learning (e.g., face recognition [5], recommendation systems [6]). In many modern applications, data sets are very large with both large number of samples \( n \) and large dimension \( p \), often with \( p \gg n \), leading to a number of unknown parameters that greatly exceeds the number of observations. This high-dimensional regime naturally calls for exploiting or assuming additional structural properties of the data to reduce the number of estimated degrees of freedom. The most popular examples of structures utilized in covariance estimation include linear models such as Toeplitz [7–16], group symmetric [17], sparse [18–20], low rank [21–23] and many others. Non-linear structures are also quite common in engineering applications, among them are the Kronecker product model [24–27], linear inverse covariance structures such as graphical models, [28, 29] and others.

In this paper we focus on the Kronecker Product (KP) structure, which has recently become a popular model for a variety of applications, such as MIMO wireless communications [30], geostatistics [31], genomics [32], multi–task learning [33], face recognition [5], recommendation systems [6], collaborative filtering [34] and more. KP model assumes a \( pq \times pq \) covariance matrix \( \Theta_0 \) to be the Kronecker product of two lower dimensional square matrices, which is denoted by \( \Theta_0 = P \otimes Q \), where \( P \) and \( Q \) are \( p \times p \) and \( q \times q \) dimensional positive definite matrices, respectively. Given \( \Theta_0 \), its factors \( P \) and \( Q \) can only be determined up to a positive scalar. This natural ambiguity is usually treated by fixing scaling of one of the factors as we do below.

Consider the Gaussian setting and assume we are given \( n \) independent and identically distributed (i.i.d.) \( pq \) dimensional real vector measurements \( x_i \sim x, \ i = 1, \ldots, n \), where

\[ x \sim N(\mu, \Theta). \]
Assume the mean $\mu$ is known, then if the number of samples is not less than the ambient dimension, $n \geq pq$, the Maximum Likelihood Estimator (MLE) of the covariance parameter exists and coincides with the Sample Covariance Matrix (SCM)

$$
S = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T.
$$

When the prior knowledge suggests that the true covariance matrix $\Theta_0$ is of KP structure, it is usually more convenient to cut $x$ into $q$ columns of height $p$ each to obtain a so called matrix normal random variable $X$, [25, 27, 35]. Following [35], we denote this law by

$$
X \sim MN(M, P \otimes Q),
$$

where $M$ is obtained from $\mu$ by the same reshaping procedure. Assume we are given $n$ i.i.d. matrix samples $X_i \sim X$, $i = 1, \ldots, n$ as in (3), and want to estimate the covariance matrix factors $P$ and $Q$. Here the MLE solution is no longer given by an explicit formula as in (2), moreover, the resulting optimization program is non-convex due to the constraint. Luckily, there exists an alternating optimization approach, which is usually adopted [25, 26, 36, 37]. This algorithm is often referred to as the Flip-Flop (FF) due to the symmetric updates of the estimates of $P$ and $Q$ it produces. Below we show that the obtained constrained program becomes convex under a specific change of the metric over the set of positive definite matrices, [38, 39], naturally explaining the convergence of the FF and significantly helping to further explore the optimization at hand. We refer to this iterative algorithm as Gaussian FF (GFF) to distinguish it from another FF scheme introduced below.

In many real world applications the underlying multivariate distribution is actually non-Gaussian and robust covariance estimation methods are required. This occurs whenever the distribution of the measurements is heavy-tailed or a small proportion of the samples exhibits outlier behavior, [40, 41]. Probably the most common extension of the Gaussian family allowing for treating heavy-tailed populations is the class of elliptically shaped distributions, [42]. Elliptical populations served the basis for defining a family of the so called covariance $M$-estimators, [41], of which we focus on Tyler’s estimator, [43]. Given $n$ i.i.d. random vectors $x_i \in \mathbb{R}^{pq}, i = 1, \ldots, n$, Tyler’s covariance matrix estimator is defined as the solution to the fixed point equation

$$
T = \frac{pq}{n} \sum_{i=1}^{n} \frac{x_i x_i^T}{x_i^T T^{-1} x_i}.
$$

When $x_i$ are Generalized Elliptically (GE) distributed [42], their shape matrix $\Theta$ is positive definite and $n > pq$, Tyler’s estimator exists with probability one and is a consistent estimator of $\Theta$ up to a positive scaling factor. The GE family includes as particular cases generalized Gaussian distribution, compound Gaussian, elliptical and many others [42]. Therefore, it has been successfully used to replace the SCM in many applications such as anomaly detection in wireless sensor networks [44], antenna array processing [45] and radar detection [8, 46–48]. It has been recently demonstrated that Tyler’s estimator can be actually viewed as an MLE of a certain spherical distribution, [38, 49, 50]. This discovery provided a very useful natural optimization framework characterizing Tyler’s estimator and allowed for defining its structured analogs, [15–17]. Similarly to the Gaussian scenario, when the number of samples is scarce compared to the dimension, the way to improve the performance is to assume structure. We rely on the MLE definition of Tyler’s estimator to impose KP structure on it and obtain a Robust Flip-Flop (RFF) analog of the GFF algorithm. As in the Gaussian case, the involved optimization becomes convex after a specific change of the underlying metric, which we extensively use in our investigation of the existence, uniqueness and convergence properties.

In both Gaussian and robust cases, one of the central questions in high-dimensional environment is what is the minimal number of samples guarantying the existence and uniqueness of the corresponding covariance MLE. As we have already mentioned, in the unconstrained Gaussian MLE it is known that
\( n = pq \) samples is enough to guarantee existence and uniqueness almost surely when the mean is known and \( n = pq + 1 \), when the mean is unknown. This number is, of course, enough in the constrained case as well, however, one would expect that this threshold can be reduced due to the decrease in the estimated number of parameters. Different necessary and sufficient conditions on the number of samples in Gaussian scenario were proposed by a large number of works, see [25, 26, 37, 51, 52] and references therein. In particular, in [25] it was claimed that the number of samples needed to guarantee both existence and uniqueness of the GFF solution in the unknown mean case equals \( \max \left[ \frac{p}{q}, \frac{q}{p} \right] + 1 \). Later, [51] showed that in fact \( \max[p, q] + 1 \) matrix valued measurements are required to guarantee uniqueness assuming the estimator exists. In [52] the authors showed by a few simple counterexamples that both results from [25] and [51] are not correct. Instead, they claimed, that “As yet, there do not seem to be existence results for the robust one. Notations Section VII - for the open cone of positive determinant by \( |\cdot| \). The boundary of the set \( X \) denotes its span and \( \partial X \) denotes its natural structure of a Euclidean space \( \mathbb{R}^p \) inherits from \( \mathcal{M}_{p \times p} \) the natural structure of a Euclidean space with the Frobenius norm. \( \mathcal{M}_{p \times q} \) denotes the Euclidean space of real \( p \times q \) matrices, written as upper case bolds \( X \). \( S(\mathbb{R}^p) \) stands for the linear space of symmetric \( p \times p \) matrices and \( \mathcal{P}(\mathbb{R}^p) \subset S(\mathbb{R}^p) \) - for the open cone of positive definite matrices inside it. Note that \( S(\mathbb{R}^p) \) inherits from \( \mathcal{M}_{p \times p} \) the natural structure of a Euclidean space with the Frobenius norm. \( \mathcal{I} \) denotes the identity matrix of an appropriate dimension. For two spaces \( \mathbb{R}^p \) and \( \mathbb{R}^q \), \( \mathbb{R}^p \otimes \mathbb{R}^q \) denotes their tensor product space and for two matrices \( P \in \mathcal{P}(\mathbb{R}^p) \), \( Q \in \mathcal{P}(\mathbb{R}^q) \), \( P \otimes Q \) denotes their Kronecker product. The spectral norm of matrix \( P \in \mathcal{P}(\mathbb{R}^p) \) is denoted by \( \|P\|_2 \) and its determinant by \( |P| \). The boundary of the set \( X \) is denoted by \( \partial X \). For two sets \( X \) and \( Y \), \( X \times Y \) denotes their Cartesian (direct) product. Given a subset \( X \) of a linear space, \( \langle X \rangle \) denotes its span and \( |X| \) - its cardinality. We use standard abbreviation a.s. to denote the almost sure convergence when the measure can be inferred from the context.

II. GAUSSIAN SETTINGS AND THE PROBLEM FORMULATION

Assume we are given \( n \) i.i.d. Gaussian matrix samples

\[
\mathbf{X}_1, \ldots, \mathbf{X}_n \sim \mathbf{X}, \quad \mathbf{X} \sim \mathcal{MN}(\mathbf{M}, P \otimes Q),
\]
where \( X_i \in \mathcal{M}_{p \times q}, \ P \in \mathcal{P}(\mathbb{R}^p) \) and \( Q \in \mathcal{P}(\mathbb{R}^q) \). Denote \( X = \{X_1, \ldots, X_n\} \), then, up to an additive constant and scaling, the negative log-likelihood reads as
\[
\tilde{f}_N(M, P \otimes Q; X) = \frac{1}{n} \sum_{i=1}^{n} \text{Tr} \left( P^{-1}(X_i - M)Q^{-1}(X_i - M)^T \right) + \log |P \otimes Q|, \tag{6}
\]
and is defined over the set \( \mathcal{M}_N = \{P \otimes Q \mid P \in \mathcal{P}(\mathbb{R}^p), \ Q \in \mathcal{P}(\mathbb{R}^q)\} \subset \mathcal{P}(\mathcal{M}_{p \times q}). \tag{7} \)

Here matrix \( P \otimes Q \) is identified with the positive operator \( P \otimes Q : \mathcal{M}_{p \times q} \to \mathcal{M}_{p \times q} \) acting by the rule \( X \to PXQ \). The scalar product on \( \mathcal{M}_{p \times q} \) is given by \( (A, B) = \text{Tr}(AB^T) \). Note that \( \mathcal{M}_N \) can be identified with the set
\[
\mathcal{M}_N \cong \{(P, Q) \mid \|P\|_2 = 1\} \subset \mathcal{P}(\mathbb{R}^p) \times \mathcal{P}(\mathbb{R}^q), \tag{8}
\]
where the specific normalization can be chosen arbitrarily.

**Remark 1.** Below we assume the following notational convention: when the set \( \mathcal{M}_N \) is viewed as a subspace of \( \mathcal{P}(\mathcal{M}_{p \times q}) \) as in (7), the arguments of the negative log-likelihood are written as \( \tilde{f}_N(M, P \otimes Q; X) \), however, when \( \mathcal{M}_N \) is identified with a subset of \( \mathcal{P}(\mathbb{R}^p) \times \mathcal{P}(\mathbb{R}^q) \) defined by (8), we write the arguments as \( \tilde{f}_N(M, P, Q; X) \), with
\[
\tilde{f}_N(M, P, Q; X) = \tilde{f}_N(M \otimes P, Q; X). \tag{9}
\]

Below we use the same rule for other similar functions and the specific representation of the underlying set can be inferred from the way arguments are written.

Identification (8) allows us to consider elements \( P \otimes Q \in \mathcal{M}_N \) (under proper normalization \( P \otimes Q = \left( \frac{1}{\|P\|_2} P \right) \otimes (\|P\|_2 Q) \) if needed) as pairs \( (P, Q) \) and endows \( \mathcal{M}_N \) with a smooth manifold structure making \( \tilde{f}_N \) a smooth function over it. The covariance MLE under the KP constraints can now be written as a solution to the following program
\[
\min_{M \in \mathcal{M}_{p \times q}, \ (P, Q) \in \mathcal{M}_N} \tilde{f}_N(M, P, Q; X). \tag{10}
\]
As in the unconstrained normal case, this program decouples into minimization w.r.t. (with respect to) the unknown mean \( M \), yielding
\[
\hat{M} = \frac{1}{n} \sum_{i=1}^{n} X_i, \tag{11}
\]
and minimization w.r.t. to \( P \) and \( Q \). Note that \( \log |P \otimes Q| = q \log |P| + p \log |Q| \) and denote \( Y_i = X_i - \hat{M} \), then the first-order optimal conditions for \( P \) and \( Q \) read as
\[
\begin{cases}
P = \frac{1}{qn} \sum_{i=1}^{n} Y_i Q^{-1} Y_i^T, \\
Q = \frac{1}{pn} \sum_{i=1}^{n} Y_i^T P^{-1} Y_i.
\end{cases} \tag{12}
\]

There does not exist a closed form analytic solution to (12) and it is usually solved via the so-called Flip-Flop (FF) iterative scheme, [25], which we call the Gaussian FF (GFF). The GFF algorithm works as follows. Starting from an initial guess \( (P_0, Q_0) \in \mathcal{M}_N \) for \( (P, Q) \), we plug them into the right-hand
Existence measurements one needs to guarantee existence, uniqueness and convergence almost surely. control is the number of samples we require, therefore, below we focus on the question how many existence and uniqueness and for the convergence of the GFF procedure. The only parameter under our scheme for its calculation, our next goal is to determine the necessary and sufficient conditions for its at hand.

III. Existence, Uniqueness and Convergence: State of the Art

Having derived the G-CARMEL (Gaussian KRonecker product MLE) solution and obtained an iterative scheme for its calculation, our next goal is to determine the necessary and sufficient conditions for its existence and uniqueness and for the convergence of the GFF procedure. The only parameter under our control is the number of samples we require, therefore, below we focus on the question how many measurements one needs to guarantee existence, uniqueness and convergence almost surely.

- **Existence.** We start from the sufficient conditions. It was claimed in [25] that \( \max \left[ \frac{p}{q}, \frac{q}{p} \right] + 1 \) samples are needed for the existence and uniqueness of the MLE solution in the Gaussian case. However, it was later shown by a counterexample in [52] that uniqueness does not follow from this condition. In addition, the authors of [52] write that "Moreover, it is not known whether it [this condition] guarantees existence, because it is not sufficient to show that all updates of the FF algorithm have full rank as is done in [25]. It could still happen that the sequence of updates converges (after infinitely many steps) to a Kronecker product that does not have a full rank with the likelihood converging to its supremum.". It is also claimed in [52] that not less than \( pq + 1 \) samples are required to ensure the existence a.s. This number of samples coincides with that needed in the unconstrained case and does not explore the KP structure. Finally, the authors of [52] conclude that nothing can be said regarding the existence if the number of samples lies inside the interval \( n \in \left[ \max \left[ \frac{p}{q}, \frac{q}{p} \right] + 1, pq \right] \).

The necessary conditions were also treated in [25], where the author claims that if the estimator exists, then \( n \geq \max \left[ \frac{p}{q}, \frac{q}{p} \right] + 1. \) This is clearly true, since if the number of samples is less than this threshold, at least one of the right-hand sides in (12) is rank deficient and cannot be invertible.

- **Uniqueness.** As summarized in [52], the author of [25] claims that the G-CARMEL is unique whenever \( n \geq \max \left[ \frac{p}{q}, \frac{q}{p} \right] + 1. \) Later, the authors of [51] stated that indeed \( n \geq \max[p, q] + 1 \) is needed to ensure uniqueness. Here again, [52] succeeded to find a counterexample showing that both these bounds do not guaranty uniqueness. Moreover, in [52] it is described what are the exact parts of the proofs which seem to contain mistakes, however, the correct lower bounds on the number of required samples are not provided. In fact, to the best of our knowledge, tight sufficient conditions for the uniqueness have not been reported so far.

- **Convergence of the Gaussian Flip-Flop Algorithm.** The last question regarding the G-CARMEL we focus on is the convergence of the GFF iterative scheme. In [25] the author establishes the convergence of the GFF technique empirically and claims that if the limiting points of the sequences \( \{\hat{Q}_k\} \) and \( \{\hat{P}_k\} \) do not depend on the initial point and an additional condition on the second derivatives of the objective is satisfied at the limiting points, then these limits provide the G-CARMEL solution. If such limiting points are not uniquely determined, but rather depend on the initial guess, they must provide a
local extrema of the likelihood function. Unfortunately, this empirical approach can hardly be applied in practice and does not provide a strict criterion for the convergence of the GFF. The authors of [36, 37] claim that when the number of samples is $n \geq pq + 1$, the GFF is guarantied to converge, however they doubt if it really converges to the MLE, since the “parameter space of $(p, q)$-separable covariance matrices is not convex”. They emphasize that for some values of $n$, the algorithm can converge to many different estimates, depending on the starting value. Finally, they conjecture that for $n$ large enough “the limit point of the GFF can safely be regarded as the unique MLE” without proving this statement.

In [26] theoretical asymptotic properties of the GFF algorithm are considered and the algorithm’s performance for small sample sizes is investigated with a simulation study.

The main contributions of our paper consist in proving tight sufficient and necessary conditions for a.s. existence and uniqueness of the G-CARMEL estimate and showing that the sufficient conditions also imply convergence of the GFF iterations to the unique solution starting from any initial guess.

IV. THE MAIN RESULTS AND ARGUMENTS

In this section we state our main result in the normal case, give the intuition behind the proof argument and demonstrate our technique on a simple example in a low dimension.

A. The Main Statement

**Theorem 1.** Assume $X = \{X_1, \ldots, X_n\}$ are independently sampled from a continuous distribution, $p \times q$ matrices and consider the problem of minimizing $\tilde{f}_N(M, P; X)$ over $\mathcal{M}_{p \times q} \times \mathcal{M}_N$, then

1) if $n < \max \left\{ \frac{p}{q}, \frac{q}{p} \right\} + 1$, there is no unique minimum,

2) if $n > \frac{p}{q} + \frac{q}{p} + 1$, there is a unique minimum a.s.,

3) if $n > \frac{p}{q} + \frac{q}{p} + 1$, the GFF converges starting from any point of $\mathcal{M}_N$ to this unique minimum a.s.

**Proof.** This is a direct Corollary of Theorem 4 from Section VI-G.

**Remark 2.** The statement of the theorem is valid for any continuous distribution and not only for the normal one. Indeed, the theorem does not claim anything about the statistical properties properties of the estimator (e.g. consistency), but rather only treats the questions of existence and uniqueness of the minimum.

**Remark 3.** Note the gap between items 1) and 2) containing one or two (when $p = q$) integer points which cannot be eliminated. We discuss this phenomenon below in more detail.

B. Sketch of the The Proof

In this section we discuss the main building blocks of the proof of Theorem 1 leaving the technical details to Section VI.

**Reduction to the Centered Case.** Let

$$\tilde{g}_N(P \otimes Q; X) = \tilde{f}_N(0, P \otimes Q; X),$$

then minimization of $g_N$ over $\mathcal{M}_N$ does not require optimization w.r.t. the mean parameter and the general case with the unknown expectation can be reduced to it via the following observation. Given a family $X = \{X_1, \ldots, X_n\} \subset \mathcal{M}_{p \times q}$ of $n$ random $p \times q$ matrices, there always exist another family $Y = \{Y_1, \ldots, Y_{n-1}\} \subset \mathcal{M}_{p \times q}$ of $n - 1$ random matrices such that

$$\tilde{g}_N(P \otimes Q; Y) = \tilde{f}_N(M, P \otimes Q; X).$$

Lemma 11 from Section VI shows why this is true and justifies our transition to the zero mean case. In the remainder of this section we treat the zero mean setting.
• **Necessary Conditions.** Since we require the solution to be composed of invertible matrices \( P \) and \( Q \), (12) must hold at the extremum point. Note that its right-hand side is not invertible for \( n < \max \left[ \frac{p}{q}, \frac{2}{p} \right] \), then returning to the non-centered case and compensating this by adding one sample yields item 1) of Theorem 1.

• **Sufficient Conditions.** To derive the sufficient conditions, in Section VI we change the parametrization of \( \tilde{g}_N \) by

\[
g_N(P \otimes Q; X) = \tilde{g}_N(P^{-1} \otimes Q^{-1}; X) = \frac{1}{n} \sum_{i=1}^{n} (PX, Q, X_i) - \log|P \otimes Q|,
\]

and introduce a specific metric over \( \mathcal{P}(\mathbb{R}^p \times \mathbb{R}^q) \) w.r.t. which the set \( \mathcal{M}_N \) and the function \( g_N(P \otimes Q; X) \) are convex. The sought for solution exists and is unique if and only if \( g_N \) continuously tends to \( +\infty \) on the boundary as shown in Lemma 4, in which case it is also strictly convex. Theorem 3 then demonstrates that this happens a.s. w.r.t. the distribution of \( X \) when \( n > \frac{q}{p} + \frac{q}{p} \). In the next section we demonstrate the reasoning of these claims by exploiting the case \( p = q = 2 \) in more detail.

• **Convergence of the GFF.** Suppose we are given a pair of matrices \( (P_0, Q_0) \in \mathcal{M}_N \) and use (13) to generate a sequence

\[
P_0 \quad P_1 \quad P_2 \quad \ldots
\]

\[
Q_0 \quad Q_1 \quad Q_2 \quad \ldots
\]

Here the successive iterates \( P_i, Q_{i+1}, \ldots \) are obtain by minimizing \( g_N \) w.r.t. \( Q \) when \( P_i \) is fixed and similarly by minimizing w.r.t. \( P \) when \( Q_i \) is fixed, etc. As we have mentioned in the previous paragraph, \( g_N \) is a.s. strictly convex and tends to \( +\infty \) on the boundary when \( n > \frac{q}{p} + \frac{q}{p} \). This guarantees a decrease of the target function on each iteration and the convergence of the sequence \( (P_n, Q_{n+1}) \) to the unique minimum, hence, \( (P_n, Q_n) \) converges as well.

Let us now illustrate the main arguments by a simple low dimensional example.

**C. \( p = q = 2 \) Case Study**

Assume \( \mathbb{R}^p = \mathbb{R}^q = \mathbb{R}^2 \), \( X = \{X_1, \ldots, X_n\} \subset \mathcal{M}_{2 \times 2} \) consists of \( n \) matrices and we deal with the case of zero mean. We are going to show a bit more than we have announced in the previous section, namely, we will prove that

1) If \( n = 1 \), the set of minima is non empty and forms a submanifold of dimension 3 with probability one. In particular, a minimum exists but is not unique.

2) If \( n = 2 \), there exists a polynomial \( D(X_1, X_2) \) such that

   • if \( D(X_1, X_2) \geq 0 \), there is no unique minimum of \( g_N \) over \( \mathcal{M}_N \),

   • if \( D(X_1, X_2) < 0 \), there is a unique minimum of \( g_N \) over \( \mathcal{M}_N \).

3) If \( n > 2 \), there is a unique minimum of \( g_N \) over \( \mathcal{M}_N \).

As explained in Section VI-B, the set \( \mathcal{M}_N \subset \mathcal{P}(\mathbb{R}^2 \times \mathbb{R}^2) \) is convex w.r.t. to a specific metric change. In addition, Lemma 4 demonstrates that the solution to the optimization at hand exists and is unique if and only if \( g_N \) continuously tends to \( +\infty \) on the boundary, which we use below.

1) \( n = 1 \) case. Here both equations in (12) (after replacing \( P \) and \( Q \) by their inverses) become identical to

\[
P^{-1} = \frac{1}{2} X_1 Q X_1^T.
\]

This equation defines a submanifold \( \mathcal{M}_m \subset \mathcal{M}_N \) isomorphic to \( \mathcal{P}(\mathbb{R}^2) \) containing \( Q \)-s, and, therefore, having dimension 3. A straightforward computation shows that the value of \( g_N \) is constant on \( \mathcal{M}_m \). Since \( g_N \) is convex, all points of \( \mathcal{M}_m \) are minima.
2) $n = 2$ case. It turns out that the critical question defining the behavior of the solution here is whether there exists a vector $t \in \mathbb{R}^2$ such that $X_1 t$ and $X_2 t$ are parallel. If the answer is negative, the minimum exists and is unique, otherwise, if it exists, it is not unique. As Lemma 4 item 4) shows below, such vector $t$ does not exist if and only if $g_N$ tends to $+\infty$ on the boundary of $\mathcal{M}_N$. Next we explain the reasoning in more detail and explicitly construct such $t$.

Consider a sequence $\mathcal{M}_N \ni \{(P_n, Q_n)\} \to \partial \mathcal{M}_N$, meaning that either $P_n$ tend to a singular matrix or the norm of $Q_n$ is unbounded (or both). In other words we distinguish between two cases: either $\|P_n\|_2$ and $\|Q_n\|_2$ are bounded or we may assume that in some appropriately chosen bases $\{s_1, s_2\}$ and $\{t_1, t_2\}$ for $P$ and $Q$, respectively, we have

$$P_n = \begin{pmatrix} \alpha_n & 0 \\ 0 & 1 \end{pmatrix}, \quad Q_n = \begin{pmatrix} \mu_n & 0 \\ 0 & \eta_n \end{pmatrix},$$

where $\alpha_n \leq 1$ and we assume (after swapping the eigenbasis of $Q_n$ if necessary) that $\mu_n \to +\infty$ not slower than $\eta_n$ (if $\eta_n$ is bounded this is vacuously true).

If the spectral norms are bounded the trace term of $g_N$ is bounded. Since in this case at least one of the matrices must tend to a singular one, $\log |P \otimes Q| \to -\infty$, implying $g_N \to +\infty$. In the second case we have sequence (19) and note that the logarithmic term of $g_N$ has summand tending to $-\infty$ with the rate not greater than $\log \mu_n$. Assume $X_1 t_1$ and $X_2 t_1$ are not parallel, then at least one of $X_i$ has a non zero 2,1) element. Suppose this is $X_1 = (x_{11}, x_{12})$ with $x_{21} \neq 0$, then the scalar products part of $g_N$ is not less than $\frac{1}{2} |x_{21}|^2 \mu_n$. Hence, this part of $g_N$ tends to $+\infty$ faster than the negative part and totally $g_N \to +\infty$ on the sequence at hand.

Now, suppose that there does exist a vector $t \in \mathbb{R}^2$ such that $X_1 t$ and $X_2 t$ are collinear. Normalize $t$ and form an orthonormal basis $\{t, t'\}$ in the space of $Q$, after this normalize $X_1 t$, which we denote by $s$ and complete it to an orthonormal basis $\{s, s'\}$ in the $P$’s space. In these bases each $X_i$ reads as (here we omit index $i$ in matrix elements for simplicity)

$$X_i = \begin{pmatrix} x_{11} & x_{12} \\ 0 & x_{22} \end{pmatrix}.$$  

Now define a new sequence in the chosen bases

$$P_n = \begin{pmatrix} \frac{1}{\mu_n} & 0 \\ 0 & 1 \end{pmatrix}, \quad Q_n = \begin{pmatrix} \mu_n & 0 \\ 0 & 1 \end{pmatrix},$$

with $\mu_n \to +\infty$. Then $P_n \otimes Q_n$ tends to the boundary of $\mathcal{M}_N$ and

$$\log |P_n \otimes Q_n| = 2 \log |P_n| + 2 \log |Q_n| = 2 \log |P_n Q_n| = 0,$$

and for each $X_i$,

$$(P_n X_i Q_n, X_i) = x_{12}^2 + x_{21}^2 + x_{22}^2 \frac{1}{\mu_n}.$$  

Hence, $g_N$ is bounded on the sequence $\{(P_n, Q_n)\}$ and we are done with this case.

Next we derive a condition on $X_1$ and $X_2$ telling whether such a mutual $t$ exists, which will suggest us the probability of such event. Let in the original bases $X_1$, $X_2$ and $t$ read as

$$X_1 = \begin{pmatrix} x \\ u \end{pmatrix}, \quad X_2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad t = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$  

We look for all triples $(X_1, X_2, t)$ such that $X_1 t$ and $X_2 t$ are collinear, which is equivalent to

$$\begin{vmatrix} \alpha x + \beta y & \alpha a + \beta b \\ \alpha u + \beta v & \alpha c + \beta d \end{vmatrix} = 0.$$  

(25)
The latter can be written as
\[ \begin{vmatrix} x & a \\ u & c \end{vmatrix} \alpha^2 + \left( \begin{vmatrix} x & b \\ u & d \end{vmatrix} + \begin{vmatrix} y & a \\ v & c \end{vmatrix} \right) \alpha \beta + \begin{vmatrix} y & b \\ v & d \end{vmatrix} \beta^2 = 0. \] (26)

Calculate the discriminant of this quadric
\[ D(X_1, X_2) = \left( \begin{vmatrix} x & b \\ u & d \end{vmatrix} + \begin{vmatrix} y & a \\ v & c \end{vmatrix} \right)^2 - 4 \begin{vmatrix} x & a \\ u & c \end{vmatrix} \begin{vmatrix} y & b \\ v & d \end{vmatrix}. \] (27)

Note that \( D(X_1, X_2) \) is a non-zero polynomial and there is \( t \in \mathbb{R}^2 \) with the required properties if and only if \( D \geq 0 \). If the density of the distribution of \((X_1, X_2)\) is a.s. non-zero, then clearly \( D \geq 0 \) and \( D < 0 \) both hold with non-zero probabilities.

3) \( n > 2 \) case. Here a similar computation shows that we a.s. cannot find a vector \( t \) such that \( X_i t \) are collinear, therefore the above arguments imply the existence and uniqueness of the minimum.

D. Remarks

To summarize, the answer to the existence and uniqueness question can be completely described in terms of the following indicator variable:

\[ \zeta(X) = \inf_{u \in \mathbb{R}^2 \setminus 0} \dim \left( \sum_{x_i \in X} X_i \langle u \rangle \right). \] (28)

When \( n = 1 \), \( \zeta(X) = 1 \) a.s., in case \( n > 2 \), \( \zeta(X) = 2 \) a.s., and these two situations correspond to the uniqueness and non-unique-ness. In the \( n = 2 \) case we have

\[ \zeta(X) = \begin{cases} 1, & \text{if } D \geq 0, \\ 2, & \text{if } D < 0, \end{cases} \] (29)

where both conditions hold with non-zero probability, i.e. \( \zeta(X) \) is not a constant a.s. This intuitively explains the one sample gap between necessary and sufficient conditions in Theorem 1.

In arbitrary dimension the ideas described above generalize as following. In order to guarantee the desired asymptotic behavior of the target function, our aim would be to avoid the following situation: there is a random subspace \( U \subseteq \mathbb{R}^q \) such that the dimension of \( \sum_{i=1}^n X_i U \) is less than \( \min\{n \dim U, p\} \) with non-zero probability. As the proof of Theorem 1 shows, when the number of samples satisfies the required condition, such event will a.s. not happen.

Let us focus on the gap between necessary and sufficient conditions appearing in the statement of Theorem 1

\[ I = \left[ \max \left( \frac{p}{q}, \frac{q}{p} \right) + 1, \frac{p}{q} + \frac{q}{p} + 1 \right] \] (30)

This interval contains 2 points in case \( p = q \) and only 1 point otherwise. The same argument as in the example above (note that in the example we considered the known mean case, therefore the \( n = 1 \) case considered there would correspond to \( n = 2 \) here) shows that when \( p = q \) and \( n = \max \left[ \frac{p}{q}, \frac{q}{p} \right] + 1 = 2 \in I \), there are multiple minima. Therefore, there is only one untreated integer point \( \frac{p}{q} + \frac{q}{p} + 1 \) left inside the interval \( I \). However, we do not investigate deeply the behavior of this remaining value due to the following reason. As the two dimensional example above suggests (this corresponds to the case \( n = 2 \) case in the example), in this case uniqueness and non-unique-ness happen with non-zero probabilities, making the analysis hard. Since there is only one untreated point left and the treatment involves quite non-trivial calculations, the game does not worth the candle. We believe that in general dimensions this missing point of the interval exhibits the same behavior, and both events “existence and uniqueness” and “existence and non-unique-ness” happen with non-zero probabilities.
V. ROBUST KRONECKER PRODUCT COVARIANCE ESTIMATION

A. Tyler’s Estimator

Tyler’s covariance estimator given by formula (4) can be equivalently introduced as a covariance parameter MLE of a certain spherical distribution [49, 50] as follows.

**Definition 1.** Assume \( \Theta_0 \in \mathcal{P}(\mathbb{R}^p) \), then

\[
p(x) = \frac{\Gamma(p/2)}{2 \sqrt{\pi}^p} \frac{1}{\sqrt{\|x^H \Theta_0^{-1} x\|^{p/2}}} \tag{31}
\]

is a probability density function of a vector \( x \in \mathbb{R}^p \) lying on a unit sphere. This distribution is usually referred to as the Real Angular Central Elliptical (RACE) distribution, [49], and we denote it as \( x \sim \mathcal{U}(\Theta_0) \). The matrix \( \Theta_0 \) is referred to as a shape matrix of the distribution and is a multiple of the covariance matrix of \( x \).

The RACE distribution is closely related to the class of Generalized Elliptical (GE) populations, which includes Gaussian, compound Gaussian, elliptical, skew-elliptical, RACE and other distributions, [53]. An important property of the GE family is that the shape matrix of a population does not change when the vector is divided by its Euclidean norm [42, 53]. After normalization, any GE vector becomes RACE distributed. This allows us to treat all these distributions together using Tyler’s estimator, which is the MLE of the shape matrix parameter in RACE populations and is unbiased, [49, 50].

B. Setting and Problem Formulation

In order to proceed to the robust estimation of a KP covariance matrix, we introduce the following setting. Assume we are given \( n \) i.i.d. centered real \( p \times q \)-matrix measurements \( X = \{X_1, \ldots, X_n\} \) and our goal is to determine what is the minimal number of samples \( n \) needed to ensure the existence and uniqueness of Tyler’s estimator under the KP constraint. We use the MLE formulation of Tyler’s estimator and consider the corresponding optimization program. Specifically, we search for positive definite \( P \) and \( Q \) minimizing the target

\[
\tilde{f}_E(P \otimes Q; X) = \frac{1}{pq} \log |P \otimes Q| + \frac{1}{n} \sum_{i=1}^{n} \log \left( \text{Tr} \left( P^{-1} X_i Q^{-1} X_i^T \right) \right), \tag{32}
\]

which is a robust version of the G-CARMEL estimator and is named R-CARMEL.

The target \( \tilde{f}_E \) is naturally defined over \( \mathcal{M}_\mathcal{N} \) introduced in (7) and, therefore, Remark 1 applies here and we use the same notational convention. In addition, \( f_E \) is scale invariant \( f_E(\lambda P \otimes Q; X) = f_E(P \otimes Q; X) \), hence, we rater consider \( \tilde{f}_E \) over

\[
\mathcal{M}_E = \mathcal{M}_\mathcal{N}/\{ P \otimes Q \sim \lambda P \otimes Q, \lambda > 0 \}. \tag{33}
\]

The induced map \( \mathcal{M}_\mathcal{N} \to \mathcal{M}_E \) is surjective and has no critical points. The composition \( \mathcal{P}(V) \times \mathcal{P}(U) \to \mathcal{M}_\mathcal{N} \to \mathcal{M}_E \) admits a section, thus, we may identify \( \mathcal{M}_E \) as

\[
\mathcal{M}_E \cong \{ (P, Q) \mid \|P\|_2 = \|Q\|_2 = 1 \} \subset \mathcal{P}(\mathbb{R}^p) \times \mathcal{P}(\mathbb{R}^q), \tag{34}
\]

which provides \( \mathcal{M}_E \) with a smooth manifold structure. The reason we still use \( \mathcal{M}_\mathcal{N} \) is the metric it possesses, whereas we cannot provide \( \mathcal{M}_E \) with a similar metric. Below we demonstrate that the same changes of parametrization and metric, as was used in the Gaussian case, make \( \tilde{f}_E \) convex and simplify the treatment significantly. On the other hand, we need \( \mathcal{M}_E \) when we talk about uniqueness of the extremum, since there is no uniqueness of the minimum of \( \tilde{f}_E \) over \( \mathcal{M}_\mathcal{N} \) due to the scaling ambiguity.
Minimization of \( \tilde{f}_E \) w.r.t. \( P \) and \( Q \) yields a system

\[
\begin{align*}
P &= \frac{1}{qn} \sum_{i=1}^{n} \frac{X_i Q^{-1} X_i^T}{\text{Tr} \left( P^{-1} X_i Q^{-1} X_i^T \right)}, \\
Q &= \frac{1}{pn} \sum_{i=1}^{n} \frac{X_i^T P^{-1} X_i}{\text{Tr} \left( Q^{-1} X_i P^{-1} X_i^T \right)}.
\end{align*}
\]

(35)

Similarly to the Gaussian case, there does not exist a closed form solution to this system, and an iterative solution is required, which we call the Robust Flip-Flop (RFF). It is also a descent algorithm and converges starting from any initial point due to a similar reasoning. If one wants to remain inside the set \( \mathcal{M}_E \) on each iteration, he has to normalize the iterates on each step

\[
\begin{align*}
\tilde{P}_{j+1} &= \frac{1}{qn} \sum_{i=1}^{n} \frac{X_i Q^{-1}_j X_i^T}{\text{Tr} \left( P^{-1}_j X_i Q^{-1}_j X_i^T \right)}, \\
\tilde{Q}_{j+1} &= \frac{1}{pn} \sum_{i=1}^{n} \frac{X_i^T P^{-1}_j X_i}{\text{Tr} \left( Q^{-1}_j X_i P^{-1}_j X_i^T \right)}, \\
P_{i+1} &= \frac{\tilde{P}_{j+1}}{\|\tilde{P}_{j+1}\|_2}, \quad Q_{i+1} = \frac{\tilde{Q}_{j+1}}{\|\tilde{Q}_{j+1}\|_2}.
\end{align*}
\]

(36)

The above reasoning regarding the scaling invariance of the solution explains that when the solution exists is unique, such normalization does not affect the convergence.

C. Main Results

In the robust setting described above, a more intuitive result concerning R-CARMEL and the RFF can be obtained.

**Theorem 2.** Assume \( X = \{X_1, \ldots, X_n\} \) are independently sampled from a continuous distribution \( p \times q \) matrices and consider the problem of minimizing \( \tilde{f}_E(P, Q; X) \) over \( \mathcal{M}_E \), then

1) if \( n < \max \left[ \frac{p}{q}, \frac{q}{p} \right] \), there is no unique minimum,

2) if \( n > \max \left[ \frac{p}{q}, \frac{q}{p} \right] \), there is a unique minimum a.s.

3) if \( n > \max \left[ \frac{p}{q}, \frac{q}{p} \right] \), the normalized RFF scheme (36) converges starting from any point of \( \mathcal{M}_E \) to this unique minimum a.s.

**Proof.** The proof can be found in the Appendix. \( \square \)

A few points are in place here.

**Remark 4.** Note that unlike the Gaussian case treated before, here the mean is assumed be known. This assumption is made due to the fact that otherwise the obtained optimization program is now well-defined. This happens due to the fact that the density (31) is concentrated on a sphere and is continuous. The known mean assumption naturally leads the the reduction of the corresponding numbers of samples by one.

**Remark 5.** Note that in the robust case, the gap between items 1) and 2) consists of one point only, which distinguishes this case from the Gaussian one. The robust result clearly provides a sharper threshold between the mode of existence and uniqueness of the MLE, and the mode where it does not exist at all.

VI. PROOF OF THEOREM 1

This section treats the Gaussian scenario and utilized a few useful concepts and techniques from commutative algebra. Therefore, for convenience we transition to a more general treatment of linear spaces, their tensor products, operators over them and a few more related notions. For this purpose the next section introduces some additional notations.
A. Additional Notations

Abstract vector spaces are denoted by capitals $V$ and are assumed to be real Euclidean spaces, their dual spaces are denoted by $V^*$. Note that the scalar product induces a canonical isomorphism $V \cong V^*$. If we identify $V$ with the space of columns $\mathbb{R}^p$ in some orthonormal basis, then $V^*$ may be identified with the space of rows $\mathbb{R}^p$. Then the dual basis on $V^*$ is also orthonormal and the isomorphism between $V$ and $V^*$ is given by the transpose map. Scalar products are denoted by $(\cdot, \cdot)$ and the corresponding norms by $\|\cdot\|$. The spectral norm of operators is denoted by $\|\cdot\|_2$. For an operator $A$, its adjoint is denoted by $A^*$.

We may naturally identify $V \otimes V^*$ with $\text{End}_\mathbb{R}(V)$ via $(v \otimes \xi)(u) = \xi(u)v$, then the scalar product on $V \otimes V^*$ induces a scalar product on $\text{End}_\mathbb{R}(V)$ such that, for any operators $A$ and $B$, $(A, B) = \text{Tr}(AB^*)$. If we identify $V$ with $\mathbb{R}^p$, then $\text{End}_\mathbb{R}(V)$ is identified with $M_{p \times p}(\mathbb{R})$, $A^*$ becomes $A^T$, and $(A, B) = \text{Tr}(AB^T)$. Given two Euclidean spaces $V$ and $U$, the scalar products on $V$ and $U$ induce one on their tensor product $V \otimes U$.

For any topological space $X$, we denote its one-point compactification by $\hat{X}$, that is, $\hat{X} = X \sqcup \{\infty\}$ with the base of neighborhoods of $\infty$ consisting of the sets $X \setminus K \sqcup \{\infty\}$ for all compact $K \subseteq X$. For a non-compact topological space $X$, we denote $\{\infty\}$ by $\partial X$. As an exception, for the real line $\mathbb{R}$, $\hat{\mathbb{R}}$ will be a two point compactification, that is, $\hat{\mathbb{R}} = \{-\infty\} \sqcup \mathbb{R} \sqcup \{+\infty\}$ endowed with the usual topology making it homeomorphic to the closed unit interval. Given two sequences $\omega_n$ and $\tau_n$, we write $\omega_n \asymp \tau_n$ if $\frac{\omega_n}{\tau_n} \to 1$.

In this section we shall treat the negative log-likelihoods as functions of $P^{-1}$ and $Q^{-1}$ (as we already did when discussing the example in Section IV-C). We do so to simplify the calculations and note that this change does affect the existence and uniqueness of extrema at hand. Accordingly, we denote

$$f_N(M, P \otimes Q; X) = \tilde{f}_N(M, P^{-1} \otimes Q^{-1}; X) = \frac{1}{n} \sum_{i=1}^n \text{Tr}(P(X_i - M)Q(X_i - M)^T) - \log|P \otimes Q|, \quad (37)$$

$$g_N(P \otimes Q; X) = f_N(0, P \otimes Q; X) = \frac{1}{n} \sum_{i=1}^n \text{Tr}(PX_iQX_i) - \log|P \otimes Q|. \quad (38)$$

B. Metric on $\mathcal{P}(V)$

We endow $\mathcal{P}(V)$ with a Riemannian metric, whose geodesics connecting any two points $P, R \in \mathcal{P}(V)$ is given by

$$\gamma_t(P, R) = P^\frac{1}{2} \left( P^{-\frac{1}{2}}RP^{-\frac{1}{2}} \right)^t P^\frac{1}{2}, \quad 0 \leq t \leq 1. \quad (39)$$

Due to the limited space we omit discussion of this metric and its properties. For more details on the relation of this metric to the MLE problems at hand consult [38, 39, 54, 55] and references therein. If we allow $t$ to run over $\mathbb{R}$, then we call the obtained curve an extended geodesic.

**Fact 1.** A direct computation shows that this metric is invariant under inversion. In addition we note that the determinant function is linear w.r.t. this metric.

**Lemma 1.** Let $V$ be a vector space, $x \in V$, $S \in \mathcal{P}(V)$ and

$$\varphi_x(t) = (S^tx, x), \quad t \in \mathbb{R}, \quad (40)$$

then its second derivative reads as

$$\varphi''_x(t) = (\ln^2(S)S^tx, x) = \left\| \ln(S)S^tx \right\|^2, \quad (41)$$

in particular, $\varphi_x(t)$ is convex. In addition, the following are equivalent

1) $\varphi''_x(t) = 0$ for some $t \in \mathbb{R}$,
2) \( \varphi''_x(t) \equiv 0 \) for all \( t \in \mathbb{R} \),

3) \( Sx = x \).

Therefore, if \( \varphi_x \) is linear in a neighborhood of some \( t_0 \), then \( \varphi_x \) is constant on \( \mathbb{R} \).

**Proof.** Since \( (S^t)' = \ln(S)S^t \) and \( \ln(S) \) commutes with any power of \( S \), we get (41) and the convexity follows.

Note that 2) implies 1) and, therefore, it is enough to show 1) \( \rightarrow 3) \rightarrow 2) \). If \( Sx = x \), then for any real \( t \), \( S^tx = x \). Therefore, \( \varphi_x \) is constant on \( \mathbb{R} \) and we get 3) \( \rightarrow 2) \). Now let \( \varphi''_x(t) = 0 \), then \( \ln(S)S^tx = 0 \). Since \( S \) is invertible, so is \( S^{\frac{1}{2}} \) and \( \ln(S)x = 0 \) or, equivalently, \( x \) is an eigenvector of \( S \) with eigenvalue 1. Finally, the last claim follows from the fact that \( \varphi_x(t) = \|S^{\frac{1}{2}}x\|^2 \geq 0 \) and the only linear nonnegative function is a constant function. \( \square \)

**Lemma 2.** Let \( V \) be a vector space, \( x \in V \), and \( P, R \in \mathcal{P}(V) \), then

\[
\omega_x(t) = (v, \gamma_t(P, R)v)
\]

(42)
is convex and the following are equivalent

1) \( \omega_x \) is linear on some open subset of \( \mathbb{R} \),

2) \( \omega_x \) is constant on the whole \( \mathbb{R} \),

3) \( Px = Rx \).

**Proof.** It is an immediate corollary of Lemma 1 if we set \( S = P^{-\frac{1}{2}}RP^{-\frac{1}{2}} \) and \( x = P^{\frac{1}{2}}v \). \( \square \)

**C. Convexity of \( \mathcal{M}_N \) and \( g_N \)**

Let \( V \) and \( U \) be vector spaces, then their tensor product naturally induces a map

\[
\otimes: \mathcal{P}(V) \times \mathcal{P}(U) \rightarrow \mathcal{P}(V \otimes U),
\]

(43)
sending \( (P, Q) \) to \( P \otimes Q \). We denote the image of this map by \( \mathcal{M}_N \). The identification

\[
\mathcal{M}_N \cong \{ (P, Q) \mid \|P\|_2 = 1 \} \subset \mathcal{P}(V) \times \mathcal{P}(U),
\]

(44)
provides \( \mathcal{M}_N \) with a structure of a smooth manifold. Intuitively, this amounts to saying that fixing the norm of the first component of the KP resolves the scaling ambiguity and provides a bijective correspondence between the factors and their products. Note that the normalization in (44) is chosen arbitrarily to, and the specific choice does not affect the existence and uniqueness results. In addition, we have

**Lemma 3.** The manifold \( \mathcal{M}_N \subset \mathcal{P}(V \otimes U) \) is convex w.r.t. the geodesic metric defined in Section VI-B.

**Proof.** Since \( P \otimes Q = P \otimes I \cdot I \otimes Q \) and \( P \otimes I \) commutes with \( I \otimes Q \), we have

\[
\gamma_t(P \otimes Q, R \otimes T) = \gamma_t(P, R) \otimes \gamma_t(Q, T),
\]

(45)
where the right-hand side is in \( \mathcal{M}_N \), thus we are done. \( \square \)

**Fact 2.** For any distinct \( P, R \in \mathcal{P}(V) \), \( \gamma_t(P, Q) \rightarrow \infty \) as \( t \rightarrow \pm \infty \).

**Lemma 4.** Let \( V \) be a vector space, \( X \subset V \) - a fixed finite subset, \( P \in \mathcal{P}(V) \) and

\[
g(P; X) = \frac{1}{|X|} \sum_{x \in X} (Px, x) - \log|P|,
\]

(46)
then

1) \( g \) is convex in the Riemannian metric (39),

2) \( g \) is linear on \( \gamma_t(P, R) \) for some \( P, R \in \mathcal{P}(V) \) iff \( Px = Rx \) for all \( x \in X \),
3) if \( g \) has two minima \( P \neq R \) in \( P(V) \), then the minimum is achieved on the whole extended geodesic \( \gamma_t(P, R), t \in \mathbb{R} \).
4) let \( U \) be another vector space, \( \mathcal{M}_N \subset P(V \otimes U) \) as before, and \( \hat{g}_N: \mathcal{M}_N \to \hat{R} \) extends \( g_N \) such that \( \hat{g}_N(\infty; X) = +\infty \), then \( \hat{g}_N \) is continuous iff \( g_N \) has a unique minimum.

**Proof.** 1) \( \log |P| \) is linear by Fact 1. Lemma 2 implies that each \( (Px, x) \) is convex, therefore, so is \( g(P; X) \).

2) This follows from the 3) \( \rightarrow \) 2) implication of Lemma 2.

3) The convexity implies that the restriction of \( g \) onto \( \gamma_t(P, R), 0 \leq t \leq 1 \), is constant and, therefore, linear. Now the 1) \( \rightarrow \) 2) implication of Lemma 2 finishes the proof.

4) Let us show the sufficiency of the condition. Indeed, if \( \hat{g}_N \) is continuous, then it achieves a minimum at some interior point (that is the existence). If such minimum is not unique, then by 3), \( \hat{g}_N \) must be constant on the whole extended geodesic and cannot be continuous when approaching the boundary, since \( \hat{g}_N(\infty; X) = +\infty \).

We proceed to the necessity. Let \( S_0 \in M_N \) be the unique minimum, \( \nu = g_N(S_0; X) \) and \( R \in T_{S_0}M_N \) be any tangent vector at \( S_0 \). Let \( \gamma_t(R) \) be the geodesic starting at \( S_0 \) in direction \( R \)
\[
\gamma_0(R) = S_0 \quad \text{and} \quad \gamma'_0(R) = R.
\]

The explicit formula for \( \gamma_t \) reads as
\[
\gamma_t(R) = C_0 e^{tC_0^{-1}RC_0^{-1}} C_0,
\]
where \( C_0 = S_0 \frac{1}{2} \) is the unique symmetric square root. In particular, \( \gamma_{\lambda t}(R) = \gamma_t(\lambda R) \) for any \( \lambda > 0 \). Set \( \delta_t(R) = g_N(\gamma_t(R); X) \). We claim, that
\[
\sigma = \min_{|R|=1} \delta'_1(R) > 0.
\]
Indeed, suppose \( \delta'_1(R_0) = 0 \) for some \( R_0 \), then since \( \delta_t(R_0) \) is convex and \( t = 0 \) is a minimum, \( \delta_t(R_0) \) is constant for \( 0 \leq t \leq 1 \). Thus, the minimum is not unique, which is a contradiction.

Denote by \( B^0_t \) the centered open ball of radius \( t \) in \( T_{S_0}M_N \) and \( B_t \) - its closure, then \( \gamma_1(B^0_t) \) is a family of open neighborhoods of \( S_0 \) with compact closure \( K_t = \gamma_1(B_t) \). Thus, we need to show that \( \inf_{P \otimes Q \in K_t} g_N(P \otimes Q; X) \to +\infty \) as \( t \to +\infty \). Indeed,
\[
\inf_{P \otimes Q \in K_t} g_N(P \otimes Q; X) = \inf_{R \in T_{S_0}M_N, \|R\| \geq t} g_N(\gamma_1(R); X) \geq g_N \left( \gamma_{\|R\|} \left( \begin{array}{c} R \\ \|R\| \end{array} \right); X \right) \geq \sigma(\|R\| - 1) + \nu
\]
\[
\geq \sigma(t - 1) + \nu,
\]
where in the last line \( R \in T_{S_0}M_N \) is any matrix of norm at least \( t \).

---

**D. The set of “bad” samples**

Depending on the set \( X \), \( g_N \) may happen to be not strictly convex on \( M_N \), or equivalently, \( \hat{g}_N \) is not necessarily continuous. In the section we discuss when this situation occurs and compute the measure of the set of samples making \( \hat{g}_N \) discontinuous.

For a vector space \( V \) of dimension \( p \) and \( d, s \in \mathbb{N} \) define
\[
G_{d,s}(V) = \{ (v_1, \ldots, v_s) \in V^s \mid \dim(\langle v_1, \ldots, v_s \rangle) = d \} \subseteq V^s
\]
to be all the \( s \)-tuples of vectors in \( V \), spanning subspaces of dimension \( d \). \( G_{d,s}(V) \) is a smooth manifold of dimension \( (p + s - d)d \). Note also that \( G_{d,d}(V) \subseteq V^d \) is an open subset, moreover, if we represent \( V^d = V \otimes \mathbb{R}^d \), we get an action of \( \text{GL}_d(\mathbb{R}) \) on \( V^d \), which restricts correctly onto \( G_{d,d}(V) \) and is free.
Before giving a precise statement about what Diagram 1 displays, let us provide an intuitive explanation of it. The operator analog of a $p \times q$ matrix is an element of $\text{Hom}(U, V)$, thus our $k$ matrix measurements in $X$ together represent an element of $\text{Hom}(U, V)^k$. We next take $d$ linearly independent vectors in a $q$ dimensional $U$ and apply all the elements of our $k$-tuple to them - this provides us with $dk$ vectors in a $p$ dimensional $V$. If we now take all the pairs of such a $k$-tuple of operators and $d$-tuples of vectors in $U$, and consider the described action of the former on the latter, which we call $\Psi$, we get the first line on the diagram at hand. Let us now consider all sets of $dk$ vectors inside $V$ spanning a subspace of dimension $r$ and take their preimage under $\Psi$, we get a subset $Z_{dkr}$ of $\text{Hom}(U, V)^k \times G_{dd}(U)$, depicted in the diagram. Finally, the leftmost arrow $\pi$ denotes the projection of this set onto $\text{Hom}(U, V)^k$. This story is made precise by the following

**Definition 2.** Let $V$ and $U$ be vector spaces of dimensions $p$ and $q$, respectively, and $k, d, r \in \mathbb{N}$ such that $d \leq q$. Consider Diagram 1, where $\Psi$ is defined as

$$\Psi : \text{Hom}(U, V)^k \times G_{dd}(U) \to V^{dk},$$

$$((\varphi_1, \ldots, \varphi_k), (u_1, \ldots, u_d)) \mapsto (\varphi_i u_j).$$

Identify $V^{dk} = V^k \otimes \mathbb{R}^d$, then $\Psi$ read as

$$\Psi((\varphi_1, \ldots, \varphi_k), (u_1, \ldots, u_d)) = \varphi_i u_j \otimes e_j,$$

and $C \in \text{GL}_d(\mathbb{R})$ acts by $I \otimes C$. Define $Z_{dkr} = \Psi^{-1}(G_{rk}(V))$ and $\pi$ to be the restriction of the projection along $G_{dd}(U)$.

**Lemma 5.** With the notations of Definition 2, for each $(u_1, \ldots, u_d) \in G_{dd}(U)$ the map

$$\Omega : \text{Hom}(U, V)^k \to V^{dk},$$

$$(\varphi_1, \ldots, \varphi_k) \mapsto (\varphi_i u_j)$$

is surjective.

**Proof.** $\Omega$ is a direct sum of $k$ maps

$$\omega : \text{Hom}(U, V) \to V^d, \quad \varphi \mapsto (\varphi u_1, \ldots, \varphi u_d).$$

Now, we choose bases in $V$ and $U$ such that $V = \mathbb{R}^p$ and $U = \mathbb{R}^q$. Then, $\text{Hom}(U, V) = \mathcal{M}_{p \times q}$ and $V^d = \mathcal{M}_{p \times d}$. Let $U = [u_1, \ldots, u_d]$, then $\omega$ reads as

$$\omega : \mathcal{M}_{p \times q}(\mathbb{R}) \to \mathcal{M}_{p \times d}(\mathbb{R}), \quad X \mapsto XU.$$ 

Since $d \leq q$ and the columns of $U$ are linearly independent, $U$ is of full rank. Hence, the map $\omega$ is surjective.

**Lemma 6.** With the notations of Definition 2,

1) $\Psi$ is surjective and $d\Psi$ is surjective at each point,
2) \( Z_{dkr} \) is a smooth manifold with
\[
\dim Z_{dkr} = pqk + qd - (p - r)(dk - r),
\]
(59)

3) a non empty fiber of \( \pi \) has dimension at least \( d^2 \).

Proof. 1) To show \( \Psi \) is surjective, let \( (u_1, \ldots, u_d) \in G_{dd}(U) \). Now it is enough to show that
\[
\Psi(\cdot, (u_1, \ldots, u_d)) : \text{Hom}(U, V)^k \to V^{dk}
\]
is surjective, which follows from Lemma 5. We proceed to the surjectivity of \( d\Psi \) at any point \( \{\varphi_i\}, \{u_j\} \).

Since \( G_{dd}(U) \subseteq U^d \) is open, identify the tangent space of \( G_{dd}(U) \) at \( \{u_j\} \) with \( U^d \), then
\[
d_{\{\varphi_i\}, \{u_j\}} \Psi : \text{Hom}(U, V)^k \times U^d \to V^{dk},
\]
(61)
\[
(\{\varphi_i^*\}, \{u_j^*\}) \mapsto (\varphi_i^* u_j + \varphi_i u_j^*).
\]
(62)

Taking \( u_j^* = 0, j = 1, \ldots, d, \) we get \( (\{\varphi_i^*\}, \{0\}) \mapsto (\varphi_i u_j) \) and the result follows from Lemma 5.

2) The fact that \( Z_{dkr} \) is a smooth manifold follows from 1) and the Implicit Function Theorem. A direct computation yields
\[
\dim Z_{dkr} = \dim A_{kd} + \dim G_{r, dk}(V) - \dim V^{dk} = pqk + qd + (p + dk - r)r - pdk
\]
(63)
\[
= pqk + qd - (p - r)(dk - r).
\]

3) \( \text{GL}_d(\mathbb{R}) \) acts freely on the fibers of \( \pi \), so the dimension of a fiber is at least \( \dim \text{GL}_d(\mathbb{R}) = d^2 \). \( \square \)

**Corollary 1.** With the notations of Lemma 6, if \( k > \frac{p}{q} + \frac{q}{p} \) and \( r \leq d\frac{p}{q} \), then \( \pi(Z_{dkr}) \) is a set of Lebesgue measure zero.

Proof. By Sard’s theorem, it is enough to show that the image of \( \pi \) consists of critical values only. So we need to show that
\[
\text{rk}(\pi) < \dim \text{Hom}(U, V)^k.
\]
(64)

Since
\[
\text{rk}(\pi) \leq \dim \pi(Z_{dkr}) - d^2,
\]
(65)
by Lemma 6 item 3), it is enough to prove that
\[
\dim \pi(Z_{dkr}) - d^2 < \dim \text{Hom}(U, V)^k.
\]
(66)
By Lemma 6 item 2), the latter is equivalent to
\[
pqk + qd - (p - r)(dk - r) < pqk,
\]
(67)
therefore, we need
\[
\frac{q - d}{p - r} + \frac{r}{d} < k, \quad \forall 1 \leq d \leq q \text{ and } 0 \leq r \leq d\frac{p}{q}.
\]
(68)
Differentiate the left hand-side by \( r \) to see that it is a strictly increasing function of \( r \), thus, it is enough to demonstrate the inequality for \( r = d\frac{p}{q} \), which is
\[
\frac{q - d}{p - d\frac{p}{q}} + \frac{p}{q} = \frac{q}{p} + \frac{p}{q} < k,
\]
(69)
and holds by the assumption. \( \square \)

**Corollary 2.** Let \( V \) and \( U \) be vector spaces of dimensions \( p \) and \( q \) respectively, and \( X \) - a finite mutually continuous family of random operators from \( U \) to \( V \) such that
\[
|X| > \frac{q}{p} + \frac{p}{q},
\]
(70)
then for each random subspace \( E \subseteq V \) we have
\[
\frac{\dim \sum_{x \in X} xE}{\dim E} > \frac{p}{q} \quad \text{a.s.}
\]  
(71)

**Proof.** Note that \( X \) is distributed on \( \text{Hom}(U, V)^k \), then
\[
\left\{ X \mid \exists E \subseteq U : \frac{\dim \sum_{x \in X} xE}{\dim E} \leq \frac{p}{q} \right\} = \bigcup_{1 \leq d \leq q} \bigcup_{0 \leq r \leq d^2} \pi(Z_{dkr}),
\]  
(72)

and the result follows from Corollary 1.
\[ \square \]

Note that in Corollary 2, we do not care whether \( p \geq q \) or \( q \geq p \). When \( p \geq q \), the inequality may fail if \( XE \) is not big enough compared to \( E \), and when \( q \geq p \), it may fail if \( E \) belongs to the kernels of all samples \( X \).

**E. Flags**

In the proof of the main theorem (see Theorem 3), we will analyze the behavior of \( g_N(P \otimes Q; X) \) when \( P \otimes Q \) tends to \( \infty \). There are many ways \( P \otimes Q \) may tend to the boundary and in order to classify all possibilities we utilize the flag machinery introduced next.

**Definition 3.** Let \( U \) be a vector space, then a flag \( F \) of length \( s \) in \( U \) is an ascending sequence of proper subspaces
\[
F = \{0 = U_0 \subsetneq U_1 \subsetneq \ldots \subsetneq U_s \mid U_s \subsetneq U\}. 
\]  
(73)

The flag is called non-trivial if \( 0 \subsetneq U_1 \subsetneq U \) and a subsequence of \( F \) is called a subflag. Let \( V \) be another vector space and \( G = \{V_i\} \) be a flag of length \( s \) in \( V \). Let \( \zeta : U \rightarrow V \) be a linear map with \( \zeta U_i \subseteq V_i \) for each \( i \leq s \), then we write \( \zeta F \subseteq G \). In addition, for all \( 0 \leq i, j \leq s \) define
\[
\Pi(F, G)_{ij} = q(dim V_j - dim V_i) - p(dim U_j - dim U_i).
\]  
(74)

If \( r = \{r_1 > \ldots > r_s > 0\} \) is a vector of strictly decreasing real numbers, we define
\[
C(F, G, r) = \sum_{i=1}^s r_i \Pi(F, G)_{i-1, i},
\]  
(75)

and note that
\[
\Pi(F, G)_{ij} + \Pi(F, G)_{jk} = \Pi(F, G)_{ik}. 
\]  
(76)

**Lemma 7.** Let \( V, U \) and \( G, F \) be as above. If \( \Pi(F, G)_{0i} > 0 \) for \( i = 1, \ldots, s \), then there exist subflags \( F' \subseteq F, G' \subseteq G \), and a subsequence \( r' \subseteq r - \text{all of length } s' \), such that
\[
\Pi(F', G')_{01} > 0, \quad \Pi(F', G')_{i-1, i} \geq 0, \quad i = 1, \ldots, s',
\]  
(77)

and \( C(F, G, r) \geq C(F', G', r') \). In particular, \( C(F, G, r) > 0 \).

**Proof.** The proof is by induction on \( s \) and the base \( s = 1 \) is the hypothesis of the lemma. Now, let \( s > 1 \) and assume the claim fails for \( F \) and \( G \), then for some \( t \leq s \)
\[
\Pi(F, G)_{i-1, i} \geq 0 \quad \forall i < t, \quad \text{and } \Pi(F, G)_{t-1, t} < 0.
\]  
(78)

Construct \( F' \) and \( G' \) from \( F \) and \( G \) by excluding \( U_{t-1} \) and \( V_{t-1} \), respectively, and \( r' \) from \( r \) by excluding \( r_{t-1} \), then
\[
C(F, G, r) = \sum_{i \neq t-1, t} r_i \Pi(F, G)_{i-1, i} + r_t \Pi(F, G)_{t-2, t-1} + \Pi(F, G)_{t-1, t} + r_t \Pi(F, G)_{t-1, t}
\]  
(79)

\[
\geq \sum_{i \neq t-1, t} r_i \Pi(F, G)_{i-1, i} + \Pi(F, G)_{t-2, t-1} + \Pi(F, G)_{t-1, t} = C(F', G', r').
\]  
\[ \square \]
F. The Known Mean Case

In this section, we prove our main result assuming the mean to be known. For this purpose we need a few auxiliary results.

Lemma 8. Let $\lambda, \mu$ and $\gamma$ be positive sequences such that $\lambda \mu = O(\log \gamma)$ and $\gamma \to +\infty$, then

$$\log \lambda^{-1} \geq \log \mu + o(\log \gamma).$$

Proof. $\lambda \mu = \alpha \log \gamma$, with $\alpha \leq \kappa$ for some constant $\kappa$. Taking logarithms yields

$$\log \lambda + \log \mu = \log \log \gamma + \log \alpha,$$

hence,

$$\log \lambda^{-1} = \log \mu - \log \log \gamma - \log \alpha \geq \log \mu - \log \log \gamma - \log \kappa = \log \mu + o(\log \gamma).$$

\[\square\]

Theorem 3. Let $V$ and $U$ be vector spaces of dimensions $p$ and $q$ respectively, $\mathcal{M}_N \subset \mathcal{P}(V \otimes U)$ as in (43) and $X \subset V \otimes U$ a finite mutually continuous family of random vectors such that $|X| > \frac{2}{p} + \frac{2}{q}$, then $g_N: \mathcal{M}_N \to \mathbb{R}$ extends to a continuous function $\hat{g}_N: \hat{\mathcal{M}}_N \to \mathbb{R}$ via $\hat{g}_N(\infty; X) = +\infty$. In particular, there exists a unique minimum of $g_N$ on $\mathcal{M}_N$.

Remark 6. Note that the statement allows the members of $X$ to be statistically dependent and does not require identical distribution. This generality is necessary when we treat the case of manually empirically centered samples below and makes application of Theorem 3 possible without additional adjustments.

Proof. By Lemma 4 item 4) it is enough to show that $g_N(P, Q; X) \to +\infty$ as $(P, Q) \to \infty$. Suppose on the contrary, there exists a sequence $(P, Q) \to \infty$ (we omit the $n$ indexing in $\{(P_n, Q_n)\}_n$ to simply notations) such that $g_N(P, Q; X) \leq \kappa$ for some constant $\kappa$. Rewrite $g_N$ as

$$g_N(P, Q; X) = \sum_{X \in X} \varphi_X(P, Q) + \psi(P, Q) = \varphi(P, Q) + \psi(P, Q),$$

where

$$\varphi_X(P, Q) = \frac{1}{|X|} (P X Q, X), \quad \psi(P, Q) = -\log|P \otimes Q|.$$  (84)

Recall that $\mathcal{M}_N$ can be identified with

$$\mathcal{M}_N \cong \{(P, Q) \mid \|P\|_2 = 1\} \subset \mathcal{P}(V) \times \mathcal{P}(U).$$

(85)

If $\|Q\|_2$ is bounded, then $(P, Q)$ tends to a singular pair (at least one of the matrices tends to a singular limit). In this case, $\varphi(P, Q) \asymp O(1)$ and $\psi(P, Q) \to +\infty$.

Now assume $\|Q\|_2 \to +\infty$, the only problem here is that $-p \log|Q|$ may tend to $-\infty$. We should show that we can compensate for this with the other summands. Let $\sigma^Q$ be the spectrum of $Q$, then it can be partitioned as $\sigma^Q = \sigma^Q_{\infty} \sqcup \sigma^Q_m$ such that

- for each $\mu \in \sigma^Q_{\infty}$, $\log \mu \asymp r_\mu \log \|Q\|_2$, where $r_\mu > 0$ is constant,
- for each $\mu \in \sigma^Q_m$, $\log \mu = o(\log \|Q\|_2)$.

Order the elements of $\sigma^Q_{\infty}$ by their rate of convergence

$$\sigma^Q_{\infty} = \sigma^Q_1 \sqcup \ldots \sqcup \sigma^Q_s,$$

where

- for each $\mu \in \sigma^Q_1$, $\mu \asymp \|Q\|_2$,
- for each $\mu, \mu' \in \sigma^Q_i$, $\lim \mu/\mu'$ is a non-zero constant,
- for any $i$, if $\mu_i \in \sigma^Q_i$, then $\mu_{i+1} = o(\mu_i)$.  (86)
For a fixed \( i \), let \( \{ \bar{K}_i \} \) be a sequence of random subspaces of \( U \) generated by the eigenvectors corresponding to \( \sigma_i \) and \( K_i \) be the limit of \( \{ \bar{K}_i \} \) (it exists after passing to an appropriate subsequence, if needed). Now \( U_k = \oplus_{j \leq k} K_j \) form a non-trivial random flag of length \( s \) in \( U \)

\[
\mathcal{F} = \{ 0 = U_0 \subseteq U_1 \subseteq U_2 \subseteq \ldots \subseteq U_s | U_s \subseteq U \}. \tag{87}
\]

Let \( \sigma^P \) be the spectrum of \( P \) and set

\[
\sigma^P_i = \{ \lambda \in \sigma^P | \lambda \mu_i = O(\log \| Q \|_2), \text{ for } \mu_i \in \sigma^Q_i \}. \tag{88}
\]

By the definition of \( \sigma^Q_i \), \( \sigma^P_i \) does not depend on the choice of \( \mu_i \in \sigma^Q_i \). Let \( \{ \bar{L}_i \} \) be the sequence of random subspaces of \( V \) generated by the eigenvectors corresponding to \( \sigma^P_i \) and \( L_i \) be the limit of \( \{ \bar{L}_i \} \) (here again, it exists after passing to an appropriate subsequence). Denote \( V_k = \oplus_{j \leq k} L_j \) and define

\[
\mathcal{G} = \{ 0 = V_0 \subseteq V_1 \subseteq \ldots \subseteq V_s | V_s \subseteq V \}, \tag{89}
\]

which is a flag of length \( s \) in \( V \). We denote the orthogonal projector in \( V \otimes U \) onto \( L_j \otimes K_i \) by \( \pi_{ij} \).

Now, there are two possibilities:

1. \( \exists X \in X : X \mathcal{F} \not\subseteq \mathcal{G} \). Let \( XU_i \not\subseteq V_i \), i.e. there is some \( j \) such that \( \log \| Q \|_2 = o(\lambda_i \mu_j) \) and \( \pi_{ij}(X) \neq 0 \), then

\[
\varphi(P, Q) \geq \varphi(X, (P, Q) = \frac{1}{|X|} (PXQ, X) \geq \frac{1}{|X|} \lambda_\mu_j \pi_{ij}(X)^2, \tag{90}
\]

and \( \psi(P, Q) = O(\log \| Q \|_2) = o(\lambda_i \mu_j) \), hence, \( g_{\mathcal{G}}(P, Q; X) \to +\infty \).

2. \( \forall X \in X : X \mathcal{F} \subseteq \mathcal{G} \). Since \( \varphi(P, Q) \not\to -\infty \), we can ignore this summand when considering the asymptotic behavior. Compute \( \psi(P, Q) = -q \log |P| - p \log |Q| \) explicitly

\[
-q \log |P| = q \sum_{i=1}^{s} \left( \dim V_i - \dim V_{i-1} \right) \log \lambda_i^{-1} - q \log \det P_{V_i^\perp}, \tag{91}
\]

\[
-p \log |Q| = -p \sum_{i=1}^{s} \left( \dim U_i - \dim U_{i-1} \right) \log \mu_i + o(\log \| Q \|_2), \tag{92}
\]

where \( \lambda_i \in \sigma^P_i \) and \( \mu_i \in \sigma^Q_i \). Note that \( -q \log \det P_{V_i^\perp} \not\to -\infty \), therefore, we may drop this summand. Since \( \mu_i \approx r_i \log \| Q \|_2 \), by Lemma 8 we obtain

\[
-\log |P \otimes Q| \gtrsim C(\mathcal{F}, \mathcal{G}, r) \log \| Q \|_2 + o(\log \| Q \|_2), \tag{93}
\]

thus, it is enough to prove that the coefficient \( C(\mathcal{F}, \mathcal{G}, r) \) is strictly positive. This would follow from see Lemma 7 if we prove that \( \Pi(\mathcal{F}, \mathcal{G})_{0i} > 0 \) for \( 1 \leq i \leq s \). Indeed,

\[
\Pi(\mathcal{F}, \mathcal{G})_{0i} = q \dim V_i - p \dim U_i = q \dim U_i \left( \frac{\dim XU_i}{\dim U_i} - \frac{p}{q} \right) \tag{94}
\]

Since \( \mathcal{F} \) is non-trivial \( \dim U_i \neq 0 \) for \( i \geq 1 \). \( |X| > \frac{q}{p} + \frac{p}{q} \), thus due to Corollary 2 the expression in brackets is a.s. strictly positive.

The proof we have just presented may be complicated to grasp due a large amount of new notations it introduces, therefore, we now explain it in an informal way. Using the same notations, let us describe the main point of using flags. Choose bases in \( U \) and \( V \) respecting the subspaces \( K_i \) and \( L_j \). In these bases all the samples \( X \in X \) has \( s \) blocks of rows and \( s \) block of columns corresponding to \( L_i \) and \( K_i \), respectively. Hence, each sample consists of \( s^2 \) blocks. Now, we can easily count the contributions of the blocks to the asymptotic of \( g_{\mathcal{G}}(P, Q; X) \).

The contributed speed of the \((i, j)\)-th block of any \( X \) is \( \lambda_i \mu_j \), up to a scalar depending on \( X \). In order to determine the asymptotic behavior of \( g_{\mathcal{G}}(P, Q; X) \) written in (83), we need to compare the negative
impact of \( \psi(P, Q) \) with the positive one of \( \varphi(P, Q) \). The highest rate negative summand appearing in \( \psi(P, Q) \) decreases with the rate of at most \( \log \|Q\|_2 \) up to a fixed scalar. If \( \lambda_i \mu_j \) tends to infinity faster than \( \log \|Q\|_2 \), then \( g_N(P, Q; X) \) would tend to \(+\infty\).

The problem appears if all the blocks corresponding to the \( \lambda_i \mu_j \) faster than \( \log \|Q\|_2 \) are zero for all \( X \in X \). Let us note that, if \((i, j)\)-th block is zero for all \( X \), then all blocks with smaller \( j \) and higher \( i \) (to the left and down of our block) have higher speed and, hence, must be zero (otherwise we are in the first situation). This precisely means that all the samples \( X \in X \) are block upper triangular, that is, they map flag \( \mathcal{F} \) into \( \mathcal{G} \).

Now we just use these observations together with Lemmas 7 and 8 to explicitly calculate the leading asymptotic term of \( \psi(P, Q) \), which, thanks to Corollary 2, turns out to be positive.

### G. The Unknown Mean Case

Let \( V \) be a vector space, \( X = \{x_1, \ldots, x_n\} \subset V \) and let \( \{e_i\} \subset \mathbb{R}^n \) be the standard basis. Define an element \( x^* \in V \otimes \mathbb{R}^n \) as \( x^* = \sum_{i=1}^{n} x_i \otimes e_i \). Then, for any \( P \in \mathcal{P}(V) \),

\[
\sum_{x \in X} (Px, x) = (P \otimes I) x^*, x^*
\]

(95)

Let now \( \hat{x} = \frac{1}{|X|} \sum_{x \in X} x \) and \( 1 = [1, \ldots, 1]^T \in \mathbb{R}^n \).

**Lemma 9.**

\[
S = \begin{pmatrix}
1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\
-\frac{1}{n} & 1 - \frac{1}{n} & \cdots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
-\frac{1}{n} & -\frac{1}{n} & \cdots & 1 - \frac{1}{n}
\end{pmatrix} \in \mathbb{R}^{n \times n}
\]

(96)

is an orthogonal projector onto a subspace of codimension 1.

**Proof.** The spectrum of \( S \) contains 0 of multiplicity 1 and the rest \( n-1 \) eigenvalues are 1-s. \( \square \)

**Lemma 10.** Let \( V \) be a vector space, \( X \subset V \) be a finite mutually continuous set of vectors, and \( P \in \mathcal{P}(V) \) - a random operator, then there is a set \( Z \subset V \) of mutually continuous vectors such that \( |Z| = |X| - 1 \) and

\[
\frac{1}{|X|} \sum_{x \in X} (P(x - \hat{x}), x - \hat{x}) = \frac{1}{|Z|} \sum_{z \in Z} (Pz, z).
\]

(97)

**Proof.** Note that \( x^* - \hat{x} \otimes 1 = (I \otimes S) x^* \). Let \( \{f_i\} \) be another orthonormal basis of \( \mathbb{R}^n \) such that \( \ker S = \langle f_n \rangle \). Compute

\[
(I \otimes S) x^* = \sum_{i=1}^{n-1} y_i \otimes f_i,
\]

(98)

and denote \( Y = \{y_1, \ldots, y_{n-1}\} \), then

\[
\sum_{x \in X} (P(x - \hat{x}), x - \hat{x}) = ((P \otimes S) x^*, x^*) = (P \otimes I y^*, y^*) = \sum_{y \in Y} (Py, y),
\]

(99)

where \( y_i \) are now centered.

Take \( z_i = \sqrt{\frac{|X| - 1}{|X|}} y_i, \ i = 1, \ldots, n - 1 \). The mutual continuity of \( z_i \) follows from the fact that the function \( V^n \to V^{n-1} \) mapping \( X \) to \( Z \) is linear and surjective. \( \square \)

**Lemma 11.** Let \( V \) and \( U \) be vector spaces, \( X \subset V \otimes U \) be a mutually continuous family of random vectors with \( |X| > 1 + \frac{q}{\mu} + \frac{q}{\nu} \), then \( f_N(\hat{X}, P, Q; X) \) as a function of \( P \) and \( Q \) extends to a continuous \( f_N: \mathcal{M}_N \to \mathbb{R} \) via \( f_N(\infty) = +\infty \), in particular, \( f_N \) has a unique minimum on \( \mathcal{M}_N \).
Proof. Applying Lemma 10, we get a mutually continuous set $Z \subset V \otimes U$ such that $|Z| = |X| - 1$ and $f_{\mathcal{N}}(X, P, Q; X) = g_{\mathcal{N}}(P, Q; Z)$, now the desired follows from Theorem 3. \hfill \Box

Theorem 4. Let $V$ and $U$ be vector spaces, $X \subset V \otimes U$ be a mutually continuous family of random vectors with $|X| > 1 + \frac{2}{p} + \frac{\xi}{p'}$, and $S = V \otimes U \times \mathcal{M}_N$, then $f_{\mathcal{N}}: S \rightarrow \mathbb{R}$ extends to a continuous function $\hat{f}_{\mathcal{N}}: \hat{S} \rightarrow \hat{\mathbb{R}}$ via $\hat{f}_{\mathcal{N}}(\infty) = +\infty$ and, in particular, there exists a unique minimum of $f_{\mathcal{N}}$ on $S$.

Proof. For a fixed pair $(P, Q)$, the $M$ minimizing $f_{\mathcal{N}}$ is the sample average, which does not depend on the values of $P$ and $Q$. Therefore, the result follows from Lemma 11 and Theorem 3. \hfill \Box

VII. PROOF OF THEOREM 2

The proof in this section is quite similar to that given in [17], thus, we made it less verbose than the proof of the previous section. For more details, please, consult [17]. Analogously to Definition 3 we introduce a notion of descending flag and note that the usage of flags in this section is different from that of Section VI.

Definition 4. Let $V$ be a real linear space, $X \subset V$ be a finite subset, $\mathcal{F} = \{V = V_0 \supseteq V_1 \supseteq \ldots \supseteq V_s \supseteq V_s \supseteq 0\}$ be a descending flag of length $s$ on $V$, define

$$\Delta(\mathcal{F}, X)_{i,j} = \dim V_i - \dim V_j - \frac{\dim V_i}{|X|} (|X \cap V_i| - |X \cap V_j|),$$

where $0 \leq i, j \leq s$. In addition, given a decreasing sequence

$$r = \{r_1 > \ldots > r_s\} \subset \mathbb{R}$$

of length $s$, define

$$S(\mathcal{F}, X, r) = \sum_{i=1}^{s} r_i \Delta(\mathcal{F}, X)_{i-1,i}.$$  

It now follows immediately from the definition that

$$\Delta(\mathcal{F}, X)_{i,j} + \Delta(\mathcal{F}, X)_{j,k} = \Delta(\mathcal{F}, X)_{i,k}, \quad i, j, k = 0, \ldots, s.$$  

Lemma 12. Let $X \subseteq V$ be a finite subset, $\mathcal{F}$ be a flag of length $s$ on $V$, $r$ be a sequence as in (101), and $\Delta(\mathcal{F}, X)_{0,i} < 0$ for all $i = 1, \ldots, s$. Then, there is a subflag $\mathcal{F}' \subseteq \mathcal{F}$ and a subsequence $r' \subseteq r$, both of length $t \leq s$ such that

$$S(\mathcal{F}, X, r) \leq S(\mathcal{F}', X, r'),$$

$$\Delta(\mathcal{F}', X)_{i-1,i} \leq 0, \quad i = 1, \ldots, t.$$  

In particular, $S(\mathcal{F}, X, r) < 0$.

Proof. The proof is by induction on $s$. For $s = 1$,

$$S(\mathcal{F}, X, r) = r_1 \Delta(\mathcal{F}, X)_{0,1} < 0,$$

Let now $s > 1$. If for all $i = 1, \ldots, s$, $\Delta(\mathcal{F}, X)_{i-1,i} \leq 0$, then we are done since $\Delta(\mathcal{F}, X)_{0,1} < 0$. Hence, we may assume that there is $i \leq s$ such that

$$\Delta(\mathcal{F}, X)_{j-1,j} \leq 0, \quad 1 \leq j < i, \quad \text{and} \quad \Delta(\mathcal{F}, X)_{i-1,i} > 0,$$

Let $\mathcal{F}'$ to be $\mathcal{F}$ without $V_i$ and $r'$ to be $r$ without $r_i$, then,

$$S(\mathcal{F}, X, r) = \sum_{j \neq i-1,i} r_j \Delta(\mathcal{F}, X)_{j-1,j} + r_{i-1} \Delta(\mathcal{F}, X)_{i-2,i-1} + r_i \Delta(\mathcal{F}, X)_{i-1,i}$$

$$\leq \sum_{j \neq i-1,i} r_j \Delta(\mathcal{F}, X)_{j-1,j} + r_{i-1} (\Delta(\mathcal{F}, X)_{i-2,i-1} + \Delta(\mathcal{F}, X)_{i-1,i}) = S(\mathcal{F}', X, r'),$$
where in the last equality we use (103). Since the length of \( \mathcal{F}' \) is less than that of \( \mathcal{F} \) and \( \Delta(\mathcal{F}', X)_{0,j} \) is either \( \Delta(\mathcal{F}, X)_{0,j-1} \) or \( \Delta(\mathcal{F}, X)_{0,j} \), thus strictly negative, the result follows by induction. \( \square \)

Let \( V \) and \( U \) be real vector spaces of dimensions \( p \) and \( q \) correspondingly. For any \( V \in V \otimes U \), denote the subspace

\[
V U^* = \{ V \xi \mid \xi \in U^* \} \subseteq V,
\]

(109)

where \( V \xi \) is the convolution along \( U \).

**Lemma 13.** Let \( V \) and \( U \) be vector spaces, and \( X \) is a family of i.i.d. continuously distributed random vectors in \( V \otimes U \), then

\[
\dim \sum_{X \in X} X U^* = \min(|X| \dim U, \dim V), \text{ a.s.}
\]

(110)

**Proof.** Choose bases in \( V \) and \( U \), then elements of \( X \) read as matrices and the space \( \sum_{X \in X} X U^* \) is spanned by the columns of all \( X \in X \). Since the elements of \( X \) are i.i.d. and continuously distributed, the matrix consisting of all columns of all \( X \)-s is of full rank. Since it has \( |X| \dim U \) columns and \( \dim V \) rows, the result follows. \( \square \)

**Corollary 3.** Let \( V \) and \( U \) be vector spaces, \( K \subsetneq V \) a proper subspace, and \( X \subset V \otimes U \) be a family of i.i.d. continuously distributed vectors, then

\[
|X \cap K \otimes U| \leq \frac{\dim K}{\dim U}, \text{ a.s.}
\]

(111)

**Proof.** Let \( Y = X \cap K \otimes U \), then \( E = \sum_{Y \in Y} Y U^* \subseteq K \) and Lemma 13 yields

\[
|Y| \dim U = \dim E \leq \dim K.
\]

(112)

\( \square \)

Similarly to the Gaussian case, below we change the parametrization

\[
f_\varepsilon(P, Q; X) = f_\varepsilon(P^{-1}, Q^{-1}; X),
\]

(113)

which does not affect the existence and uniqueness results. Partition \( f_\varepsilon(P, Q; X) \) as

\[
f_\varepsilon(P, Q; X) = -\frac{1}{p} \log|P| - \frac{1}{q} \log|Q| + \frac{1}{n} \sum_{i=1}^{n} \log \left( \text{Tr} \left( PX_i QX_i^T \right) \right) = f_P + f_Q + f_X.
\]

(114)

and consider it over \( \mathcal{M}_\varepsilon \) defined in (34), which in our new notations means that

\[
\text{Tr}(P^{-1}) = \text{Tr}(Q^{-1}) = 1.
\]

(115)

**Lemma 14.** Let \( \dim V = p \) and \( \dim U = q \), then if

\[
|X| > \frac{\max(p, q)}{\min(p, q)},
\]

(116)

\[
f_\varepsilon \to +\infty \text{ as } \mathcal{M}_\varepsilon \ni (P, Q) \to \partial \mathcal{M}_\varepsilon, \text{ a.s.}
\]

(117)

**Proof.** Assume on the contrary, there is a sequence \( T = (P, Q) \subset \mathcal{M}_\varepsilon \) (we omit indices for brevity) tending to \( \partial \mathcal{M}_\varepsilon \) and such that \( f_\varepsilon(T) \) is bounded. Note that due to (115), at least a part of eigenvalues of \( P \) and \( Q \) tend to \( +\infty \) and the others are bounded by positive constants from below. Passing to a subsequence if needed, we may assume that \( T \) converges to a point on the boundary. Below we do not mention explicitly the subsequence argument while it is assumed to be utilized if necessary.

Let \( P = \sum_{j=1}^{p} \lambda_j y_j y_j^T \) and \( Q = \sum_{i=1}^{q} \mu_i z_i z_i^T \) be the spectral decompositions of \( P \) and \( Q \) and we suppose everything to converge here. Denote the sets of eigenvalues of \( P \) and \( Q \) by \( \Lambda \) and \( \Theta \), respectively. Let \( \Lambda = \bigcup_{i=1}^{u} \Lambda_i \) and \( \rho \) be a sequence such that \( \log \lambda / \log \rho \to r_i \) whenever \( \lambda \in \Lambda_i \) and \( r_1 > \cdots > r_u > \cdots \).
In particular, \( \log \lambda \asymp r_i \log \rho \) for \( \lambda \in \Lambda_i \). Then, we define \( K_i \) to be the space generated by the limits of eigenvectors corresponding to the values in \( \Lambda_i \). Hence, \( V = \bigoplus_{i=1}^{u+1} K_i \). Now, we set \( V_i = (\bigoplus_{j=1}^{u+1} K_j)^\perp \) for all \( i = 0, \ldots, u \). Define a flag \( F \) of length \( u \) as

\[
F = \{ V \otimes U = V_0 \subset U \supset \cdots \supset V_u \otimes U \} \tag{118}
\]

and \( r = \{ r_1, \ldots, r_u \} \).

In a similar way, let \( M = \bigcup_{j=1}^{v+1} M_j \) and \( \nu \) be such a sequence that \( \log \mu/\log \nu \to t_j \) whenever \( \mu \in M_j \) and \( t_1 > \cdots > t_v > t_{v+1} = 0 \). In particular, \( \log \mu \asymp t_j \log \nu \) for \( \mu \in M_i \). The space generated by the limits of eigenvectors corresponding to \( M_i \) will be denoted by \( L_j \). Hence, \( U = \bigoplus_{j=1}^{v+1} L_j \). Now, we set \( U_i = (\bigoplus_{j=1}^{v+1} L_j)^\perp \) for all \( i = 0, \ldots, v \). Define a flag \( G \) of length \( v \) as

\[
G = \{ V \otimes U = V \otimes U_0 \supset \cdots \supset V \otimes U_v \} \tag{119}
\]

and \( t = \{ t_1, \ldots, t_v \} \).

Let \( E_{ij} = V_{i-1} \otimes U_{j-1} \subseteq V \otimes U \), then for any \( \lambda \in \Lambda_i \), \( \mu \in M_j \) and \( X \in V_{i-1} \otimes U_{j-1} \), the limit of

\[
\frac{1}{\lambda} \text{PXQ} \tag{120}
\]

exists and will be denoted by \( R_{ij}(X) \). By the definition, \( R_{ij} \) is a composition of the orthogonal projection onto \( E_{ij} \) and a positive operator on the image. Let

\[
X_{ij} = X \cap V_{i-1} \otimes U_{j-1} \setminus X \cap (V_i \otimes U_{j-1} + V_{i-1} \otimes U_j), \tag{121}
\]

then \( X = \bigcup_{i=1}^{u+1} \bigcup_{j=1}^{v+1} X_{ij} \).

We now proceed to computing the leading asymptotic terms of the summand in (114).

\[
f_P \asymp -\frac{1}{p} \sum_{i=1}^{u} \sum_{\lambda \in \Lambda_i} r_i \log \rho = -\frac{1}{p} \sum_{i=1}^{u} r_i |\Lambda_i| \log \rho = -\sum_{i=1}^{u} r_i \frac{\dim V_{i-1} - \dim V_i \log \rho}{p}. \tag{122}
\]

Similarly,

\[
f_Q \asymp -\sum_{j=1}^{v} t_j \frac{\dim U_{j-1} - \dim U_j \log \nu}{\nu}. \tag{123}
\]

Let \( X \in X_{ij} \), then for any \( \lambda \in \Lambda_i \) and \( \mu \in M_j \), we have

\[
\log(\text{PXQ}, X) \asymp \log \lambda + \log \mu + \log(\lambda^{-1} \text{PXQ}^{-1} \mu) \asymp r_i \log \rho + t_j \log \nu + \log(\text{R}_{ij}(X), X) \asymp r_i \log \rho + t_j \log \nu. \tag{124}
\]

Taking this into account, we compute

\[
f_X \asymp \frac{1}{|X|} \sum_{i=1}^{u} \sum_{j=1}^{v} \sum_{x \in X_{ij}} (r_i \log \rho + t_i \log \nu) = \frac{1}{|X|} \sum_{i=1}^{u} \sum_{j=1}^{v} |X_{ij}| (r_i \log \rho + t_i \log \nu). \tag{125}
\]

We are interested in the asymptotic of the sum \( f_P + f_Q + f_X \), whose leading term, when non-zero, can be written as

\[
f_{\infty} = f_P + f_Q + f_X \asymp A \log \rho + B \log \nu, \tag{126}
\]

where

\[
A = -\sum_{i=1}^{u} r_i \frac{\dim V_{i-1} - \dim V_i}{p} + \frac{1}{|X|} \sum_{i=1}^{u} \sum_{j=1}^{v} |X_{ij}| r_i = -\sum_{i=1}^{u} r_i \left( \frac{\dim V_{i-1} - \dim V_i}{p} - \frac{\sum_{j=1}^{v+1} |X_{ij}|}{|X|} \right) \tag{127}
\]

\[= -\sum_{i=1}^{u} r_i \left( \frac{\dim V_{i-1} \otimes U - \dim V_i \otimes U}{pq} - \frac{|X \cap V_{i-1} \otimes U| - |X \cap V_i \otimes U|}{|X|} \right) = -S(F, X, r). \]
Similar derivation yields $B = S(G, X, t)$. Thus,
\[
f_E \asymp -S(F, X, r) \log \rho - S(G, X, t) \log \nu, \tag{128}
\]
where the right-hand side is non-zero since at least one pair of eigenvalues tend to $+\infty$ due to the trace constraint (115). In addition, this implies that $\log \rho$ and $\log \nu$ tend to $+\infty$, and it remains to show the coefficients $S(\cdot)$ are both strictly negative, thus guarantying that $f_E \to +\infty$. By Lemma 12, it is enough to check that $\Delta(F, X)_{0i} < 0$ for $i = 1, \ldots, u$ and $\Delta(G, X)_{0j} < 0$ for $j = 1, \ldots, v$. We have
\[
\Delta(F, X)_{0i} = \frac{\dim V \otimes U - \dim V_i \otimes U}{pq} - \frac{|X| - |X \cap V_i \otimes U|}{|X|} = \frac{|X \cap V_i \otimes U|}{|X|} - \frac{\dim V_i}{p}, \tag{129}
\]
and we need to show
\[
\frac{|X \cap V_i \otimes U|}{|X|} < \frac{\dim V_i}{p}. \tag{130}
\]
By Corollary 3, we get
\[
\frac{|X \cap V_i \otimes U|}{|X|} \leq \frac{\dim V_i}{q|X|} < \frac{\dim V_i}{\dim V}, \tag{131}
\]
where the last inequality holds because $|X| > \frac{\max(p,q)}{\min(p,q)}$. After a similar calculation for $B$, we see that both $A$ and $B$ are negative and $f_E(T; X) \to +\infty$ as $T \to \partial M_E$. This contradicts the boundedness of $f_E(T; X)$ and finishes the proof. \hfill \Box

**Proof of Theorem 2.** Since the KP constraints and the target function are convex in the Riemannian metric we consider, the only thing we need to show is that $f_E \to +\infty$ when we approach $\partial M_E$. Lemma 14 proves that this is a.s. true when
\[
|X| > \frac{\max(p,q)}{\min(p,q)}. \tag{132}
\]
\hfill \Box

**References**


