

Disentangling Orthogonal Matrices

Teng Zhang¹

*Department of Mathematics, University of Central Florida
Orlando, Florida 32816, USA*

Amit Singer²

*Department of Mathematics and The Program in Applied and Computational Mathematics
(PACM), Princeton University
Princeton, New Jersey 08544, USA*

Abstract

Motivated by a certain molecular reconstruction methodology in cryo-electron microscopy, we consider the problem of solving a linear system with two unknown orthogonal matrices, which is a generalization of the well-known orthogonal Procrustes problem. We propose an algorithm based on a semi-definite programming (SDP) relaxation, and give a theoretical guarantee for its performance. Both theoretically and empirically, the proposed algorithm performs better than the naïve approach of solving the linear system directly without the orthogonal constraints. We also consider the generalization to linear systems with more than two unknown orthogonal matrices.

Keywords: SDP relaxation, orthogonal Procrustes problem, cryo-EM
MSC[2010] 15A24, 90C22

1. Introduction

In this paper, we consider the following problem: given known matrices $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{N \times D}$ and unknown orthogonal matrices $\mathbf{V}_1, \mathbf{V}_2 \in O(D)$, recover \mathbf{V}_1 and \mathbf{V}_2 from $\mathbf{X}_3 \in \mathbb{R}^{N \times D}$ defined by

$$\mathbf{X}_3 = \mathbf{X}_1 \mathbf{V}_1 + \mathbf{X}_2 \mathbf{V}_2. \quad (1)$$

A naïve approach would be solving (1) while dropping the constraints of orthogonality on \mathbf{V}_1 and \mathbf{V}_2 . This linear system has ND linear constraints and $2D^2$ unknown variables, therefore, this approach can recover \mathbf{V}_1 and \mathbf{V}_2 when $N \geq 2D$. The question is, can we develop an algorithm that takes the constraints of orthogonality into consideration, so that it is able to recover \mathbf{V}_1 and

¹teng.zhang@ucf.edu

²amits@math.princeton.edu

\mathbf{V}_2 when $N < 2D$, and more stably when the observation \mathbf{X}_3 is contaminated by noise?

The associated least squares problem

$$\min_{\mathbf{V}_1, \mathbf{V}_2 \in O(D)} \|\mathbf{X}_1 \mathbf{V}_1 + \mathbf{X}_2 \mathbf{V}_2 - \mathbf{X}_3\|_F^2 \quad (2)$$

can be considered as a generalization of the well-known orthogonal Procrustes problem [1]:

$$\min_{\mathbf{V} \in O(D)} \|\mathbf{X}_1 \mathbf{V} - \mathbf{X}_2\|_F^2, \quad (3)$$

with the main difference being that the minimization in (2) is over two orthogonal matrices instead of just one as in (3). Although the orthogonal Procrustes problem has a closed form solution using the singular value decomposition, problem (2) does not enjoy this property.

Still, (2) can be reformulated so that it belongs to a wider class of problems called the little Grothendieck problem [2], which again belongs to QO-OC (Quadratic Optimization under Orthogonality Constraints) considered by Nemirovski [3]. QO-OCs have been well studied and include many important problems as special cases, such as Max-Cut [4] and generalized orthogonal Procrustes [5, 6, 7]

$$\min_{\mathbf{V}_1, \dots, \mathbf{V}_n \in O(D)} \sum_{1 \leq i, j \leq n} \|\mathbf{X}_i \mathbf{V}_i - \mathbf{X}_j \mathbf{V}_j\|_F^2,$$

which has applications to areas such as psychometrics, image and shape analysis and biometric identification.

The non-commutative little Grothendieck problem [8] is defined by:

$$\min_{\mathbf{V}_1, \dots, \mathbf{V}_n \in O(D)} \sum_{i, j=1}^n \text{tr}(\mathbf{C}_{ij} \mathbf{V}_i \mathbf{V}_j^\top). \quad (4)$$

Problem (2) can be considered as a special case of (4) with $n = 3$. The argument is as follows. For convenience, we homogenize (1) by introducing a slack unitary variable $\mathbf{V}_3 \in O(D)$ and consider the augmented linear system

$$\mathbf{X}_1 \mathbf{V}_1 + \mathbf{X}_2 \mathbf{V}_2 + \mathbf{X}_3 \mathbf{V}_3 = \mathbf{0} \quad (5)$$

Clearly, if $(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3)$ is a solution to (5), then the pair $(-\mathbf{V}_1 \mathbf{V}_3^\top, -\mathbf{V}_2 \mathbf{V}_3^\top)$ is a solution to the original linear system (1). The least squares formulation corresponding to (5) is

$$\min_{\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3 \in O(D)} \|\mathbf{X}_1 \mathbf{V}_1 + \mathbf{X}_2 \mathbf{V}_2 + \mathbf{X}_3 \mathbf{V}_3\|_F^2. \quad (6)$$

Let $\mathbf{C} \in \mathbb{R}^{3D \times 3D}$ be a Hermitian matrix with the (i, j) -th $D \times D$ block given by $\mathbf{C}_{ij} = \mathbf{X}_i^\top \mathbf{X}_j$. The least squares problem (6) is equivalent to

$$\min_{\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3 \in O(D)} \sum_{i, j=1}^3 \text{tr}(\mathbf{C}_{ij} \mathbf{V}_j \mathbf{V}_i^\top), \quad (7)$$

which is the little Grothendieck problem (4) with $n = 3$.

1.1. Motivation

Our problem arises naturally in single particle reconstruction (SPR) from cryo-electron microscopy (EM), where the goal is to determine the 3D structure of a macromolecule complex from 2D projection images of identical, but randomly oriented, copies of the macromolecule. Zvi Kam [9] showed that the spherical harmonic expansion coefficients of the Fourier transform of the 3D molecule, when arranged as matrices, can be estimated from 2D projection images up to an orthogonal matrix (for each degree of spherical harmonics). Based on this observation, Bhamre et al. [10] recently proposed “Orthogonal Replacement” (OR), an orthogonal matrix retrieval procedure in which cryo-EM projection images are available for two unknown structures $\varphi^{(1)}$ and $\varphi^{(2)}$ whose difference $\varphi^{(2)} - \varphi^{(1)}$ is known. It follows from Kam’s theory that we are given the spherical harmonic expansion coefficients of $\varphi^{(1)}$ and $\varphi^{(2)}$ up to an orthogonal matrix, and their difference. Then the problem of recovering the spherical harmonic expansion coefficients of $\varphi^{(1)}$ and $\varphi^{(2)}$ is reduced to the mathematical problem (1). If (1) can be solved for smaller N , then we can reconstruct $\varphi^{(1)}$ and $\varphi^{(2)}$ with higher resolution. The cryo-EM application serves as the main motivation of this paper. We refer the reader to [10] for further details regarding the specific application to cryo-EM.

2. Algorithm and Main result

The little Grothendieck problem and QO-OCs are generally intractable, for example, it is well-known that the Max-Cut Problem is NP-hard. Many approximation algorithms have been proposed and analyzed [4, 3, 11, 12, 2, 13], and the principle of these algorithms is to apply a semi-definite programming (SDP) relaxation followed by a rounding procedure. The SDP can be solved in polynomial time (for any finite precision). Based on the same principle, we relax the problem (7) to an SDP as follows.

Let $\mathbf{H} \in \mathbb{R}^{3D \times 3D}$ be a Hermitian matrix with the (i, j) -th $D \times D$ block given by $\mathbf{H}_{ij} = \mathbf{V}_i \mathbf{V}_j^\top$, that is,

$$\mathbf{H} = \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \mathbf{V}_3 \end{pmatrix} (\mathbf{V}_1^\top, \mathbf{V}_2^\top, \mathbf{V}_3^\top).$$

Then (7) is equivalent to

$$\min_{\mathbf{H} \succcurlyeq \mathbf{0}, \mathbf{H}_{ii} = \mathbf{I}, \text{rank}(\mathbf{H}) = D} \text{tr}(\mathbf{C}\mathbf{H}),$$

where $\mathbf{H} \succcurlyeq \mathbf{0}$ denotes that \mathbf{H} is a positive semidefinite matrix. The only constraint which is non-convex is the rank constraint. Dropping it leads to the following SDP:

$$\min \text{tr}(\mathbf{C}\mathbf{H}), \text{ subject to } \mathbf{H} \succcurlyeq \mathbf{0} \text{ and } \mathbf{H}_{ii} = \mathbf{I}. \quad (8)$$

If the solution satisfies $\text{rank}(\mathbf{H}) = D$, then \mathbf{V}_1 , \mathbf{V}_2 and \mathbf{V}_3 are extracted by applying decomposition to \mathbf{H} as follows. Let $\mathbf{H} = \mathbf{U}\mathbf{U}^\top$, where

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \\ \mathbf{U}_3 \end{pmatrix} \in \mathbb{R}^{3D \times D}, \text{ and } \mathbf{U}_i \in \mathbb{R}^{D \times D}$$

then $\mathbf{V}_i = \mathbf{U}_i$, $1 \leq i \leq 3$ would be a solution.

Notice that the solution to (5) is not unique: if $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)$ satisfies (5), then for any $\mathbf{U} \in O(D)$, the triplet $(\mathbf{U}_1\mathbf{U}, \mathbf{U}_2\mathbf{U}, \mathbf{U}_3\mathbf{U})$ satisfies (5) as well. Although the solution to (5) is not unique, the solution to the original problem (2) is uniquely given by $(-\mathbf{U}_1\mathbf{U}_3^\top, -\mathbf{U}_2\mathbf{U}_3^\top)$.

When $\text{rank}(\mathbf{H}) > D$, then there does not exist $\mathbf{U} \in \mathbb{R}^{3D \times D}$ such that $\mathbf{H} = \mathbf{U}\mathbf{U}^\top$ and the linear system (1) does not have a solution. However, we could employ the rounding procedure described in [2]: Let $\mathbf{H} = \mathbf{U}\mathbf{U}^\top$, where

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \\ \mathbf{U}_3 \end{pmatrix} \in \mathbb{R}^{3D \times 3D}, \text{ and } \mathbf{U}_i \in \mathbb{R}^{D \times 3D},$$

then we generate approximate solutions by $\mathbf{V}_1 = f(-\mathbf{U}_1\mathbf{U}_3^\top)$ and $\mathbf{V}_2 = f(-\mathbf{U}_2\mathbf{U}_3^\top)$, where f is a rounding procedure to the nearest orthogonal matrix as follows. For any $\mathbf{Z} \in \mathbb{R}^{D \times D}$ with SVD decomposition $\mathbf{Z} = \mathbf{U}_\mathbf{Z}\Sigma_\mathbf{Z}\mathbf{V}_\mathbf{Z}^\top$, $f(\mathbf{Z}) = \mathbf{U}_\mathbf{Z}\mathbf{V}_\mathbf{Z}^\top = \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-\frac{1}{2}}$ [14].

2.1. Main results

The main contribution of this paper is a particular theoretical guarantee for the SDP approach to return a solution of rank D and recover \mathbf{V}_1 and \mathbf{V}_2 exactly. We start with a theorem that controls the lower bound of the objective function in (8). Throughout the paper, for any d -dimensional subspace \mathbf{L} in \mathbb{R}^D , $\mathbf{P}_\mathbf{L}$ is a projector of size $D \times d$ to the subspace.

Theorem 2.1. *For generic $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{N \times D}$ with $N \geq D+1$, $\mathbf{X}_3 = -\mathbf{X}_1 - \mathbf{X}_2$, $k \geq D$, and*

$$\mathcal{U} = \left\{ \mathbf{U} = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \\ \mathbf{U}_3 \end{pmatrix} \in \mathbb{R}^{3D \times k} : \mathbf{U}_i \in \mathbb{R}^{D \times k}, \mathbf{U}_1\mathbf{U}_1^\top = \mathbf{U}_2\mathbf{U}_2^\top = \mathbf{U}_3\mathbf{U}_3^\top = \mathbf{I}, \right\},$$

then for any $\mathbf{U} \in \mathcal{U}$,

$$\|\mathbf{X}\mathbf{U}\|_F \geq c(\mathbf{X}_1, \mathbf{X}_2) \|\mathbf{U}^\top \mathbf{P}_{\mathbf{L}^\perp}\|_F^2, \quad (9)$$

where $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) \in \mathbb{R}^{N \times 3D}$,

$$\mathbf{L}_1 = \{\mathbf{x} \in \mathbb{R}^{3D} : \mathbf{x} = (\mathbf{v}, \mathbf{v}, \mathbf{v}) \text{ for some } \mathbf{v} \in \mathbb{R}^D\},$$

and $c(\mathbf{X}_1, \mathbf{X}_2)$ is a constant depending on \mathbf{X}_1 and \mathbf{X}_2 and it is positive for generic \mathbf{X}_1 and \mathbf{X}_2 .

Based on Theorem 2.1 and $\|\mathbf{X}\mathbf{U}\|_F^2 = \text{tr}(\mathbf{C}\mathbf{H})$, this paper proves that when $N \geq D + 1$, the SDP method recovers the orthogonal matrices for generic cases, i.e., the property holds for $(\mathbf{X}_1, \mathbf{X}_2)$ that lies in a dense open subset of $\mathbb{R}^{N \times D} \times \mathbb{R}^{N \times D}$. This is formally stated next. Its part (b) shows that the SDP method is stable to noise.

Theorem 2.2. (a) For generic $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{N \times D}$ with $N \geq D + 1$, the SDP method recovers \mathbf{V}_1 and \mathbf{V}_2 exactly.

(b) Under the assumptions in (a), and suppose that the input matrices of the SDP method are $\hat{\mathbf{X}}_i$ such that $\|\hat{\mathbf{X}}_i - \mathbf{X}_i\|_F \leq \epsilon$ for $1 \leq i \leq 3$, then the SDP method recovers $\{\mathbf{V}_i\}_{i=1}^2$ approximately in the sense that the error between the recovered orthogonal matrix $\hat{\mathbf{V}}_i$ and the true orthogonal matrix \mathbf{V}_i , $\|\hat{\mathbf{V}}_i - \mathbf{V}_i\|_F$, is bounded above by $C\sqrt{\epsilon}$ for some C that does not depend on ϵ .

The result (a) shows that the SDP method successfully recovers the orthogonal matrices as long as $N \geq D + 1$, compared with the stringent requirement $N \geq 2D$ for the naïve least squares approach. The condition $N \geq D + 1$ is nearly optimal. In (1), there are ND constraints and $D(D - 1)$ variables. Hence, it is impossible to recover \mathbf{V}_1 and \mathbf{V}_2 when $N < D - 1$.

The result (b) shows that the SDP method is stable to noise in the input matrices. We remark that it might be possible to improve the stability analysis: While the current analysis gives an error of $O(\sqrt{\epsilon})$, the empirical performance usually has an error of $O(\epsilon)$, as shown in Table 2 of Section 3.

We also remark that Theorem 2.2 can be generalized to the complex case—the proof applies to the case of unitary matrices as well. For the complex case, there are $2ND$ constraints and $2D^2$ degrees of freedom. Therefore, it is impossible to recover \mathbf{V}_1 and \mathbf{V}_2 when $N < D$. Moreover, we suspect that recovery is impossible even for $N = D$, which would suggest that the sufficient condition $N \geq D + 1$ in Theorem 2.2(a) is also necessary: in fact, it is easy to verify the impossibility of recovering \mathbf{V}_1 and \mathbf{V}_2 when $N = D = 1$.

2.2. Generalization

A natural generalization of (1) is the following problem: given known matrices $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{K-1} \in \mathbb{R}^{N \times D}$ and unknown orthogonal matrices $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_{K-1} \in O(D)$, recover $\{\mathbf{V}_i\}_{i=1}^{K-1}$ from

$$\mathbf{X}_K = \sum_{i=1}^{K-1} \mathbf{X}_i \mathbf{V}_i. \quad (10)$$

For this generalized problem, the SDP method is formulated as follows. We first homogenize it to

$$\sum_{i=1}^K \mathbf{X}_i \mathbf{V}_i = \mathbf{0},$$

and let $\mathbf{H} \in \mathbb{R}^{KD \times KD}$ be a Hermitian matrix with the (i, j) -th $D \times D$ block given by $\mathbf{H}_{ij} = \mathbf{V}_i \mathbf{V}_j^\top$. Then the SDP method solves

$$\min \text{tr}(\mathbf{C}\mathbf{H}), \text{ subject to } \mathbf{H} \succcurlyeq \mathbf{0} \text{ and } \mathbf{H}_{ii} = \mathbf{I} \text{ for all } 1 \leq i \leq K, \quad (11)$$

where \mathbf{H}_{ii} represents the (i, i) -th $D \times D$ block. Then we extract the orthogonal matrices by the procedure described in Section 2.

For this generalized problem and its associated SDP approach, we have the following theoretical guarantee.

Theorem 2.3. *For generic $\{\mathbf{X}_i\}_{i=1}^{K-1} \in \mathbb{R}^{N \times D}$, if $N \geq (K-2)D + 1$, then the SDP method recovers $\{\mathbf{V}_i\}_{i=1}^{K-1}$ exactly.*

Theorem 2.2(a) can be considered as a special case of Theorem 2.3 when $K = 3$. However, for $K > 3$, the condition $N \geq (K-2)D + 1$ is not close to optimal. Since (10) has ND constraints and $D(D-1)(K-1)/2$ variables, the information-theoretic limit is $N = (D-1)(K-1)/2$. Simulations in Section 3 also show that the SDP approach empirically recovers the orthogonal matrices even when N is smaller than $(K-2)D + 1$. However, the theoretical guarantee in Theorem 2.3 is still more powerful than the least squares approach of solving $\min_{\{\mathbf{V}_i\}_{i=1}^{K-1} \in \mathbb{R}^{D \times D}} \|\sum_{i=1}^{K-1} \mathbf{X}_i \mathbf{V}_i - \mathbf{X}_K\|_F^2$, which requires $N \geq (K-1)D + 1$ to recover $\{\mathbf{V}_i\}_{i=1}^{K-1}$.

3. Numerical Experiments

In this section, we compare several methods for solving (1) and (10) on artificial data sets. The data sets are generated as follows: $\{\mathbf{X}_i\}_{i=1}^{K-1}$ are random matrices with i.i.d standard Gaussian entries $\mathcal{N}(0, 1)$, $\{\mathbf{V}_i\}_{i=1}^{K-1}$ are random orthogonal matrices (according to Haar measure) generated by QR decomposition of random matrices with i.i.d standard Gaussian entries, and \mathbf{X}_K is generated by (10).

We compare the following five methods:

1. The SDP relaxation approach (SDP) described in Section 2.
2. The naïve least squares approach (LS):

$$\min_{\{\mathbf{V}_i\}_{i=1}^{K-1}} \|\mathbf{X}_K - \sum_{i=1}^{K-1} \mathbf{X}_i \mathbf{V}_i\|_F^2$$

3. Since the convex hull of the set of orthogonal matrices is the set of matrices with operator norm not greater than one, we can strengthen the LS approach by constraining its domain (C-LS):

$$\min_{\{\mathbf{V}_i\}_{i=1}^{K-1}} \|\mathbf{X}_K - \sum_{i=1}^{K-1} \mathbf{X}_i \mathbf{V}_i\|_F^2, \text{ subject to } \|\mathbf{V}_i\| \leq 1, 1 \leq i \leq K-1$$

4. This is an approach suggested to us by Afonso Bandeira. Let us start with the case $K = 3$. If $\mathbf{V}_3 = \mathbf{V}_1 \mathbf{V}_2^\top$, then from (1), $\mathbf{X}_3 \mathbf{V}_2^\top = \mathbf{X}_1 \mathbf{V}_3 + \mathbf{X}_2$ and $\mathbf{X}_3 \mathbf{V}_1^\top = \mathbf{X}_1 + \mathbf{X}_2 \mathbf{V}_3^\top$. Then we solve the expanded least squares problem based on these three equations (LS+):

$$\min_{\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3} \|\mathbf{X}_3 - \mathbf{X}_1 \mathbf{V}_1 - \mathbf{X}_2 \mathbf{V}_2\|_F^2 + \|\mathbf{X}_3 \mathbf{V}_2^\top - \mathbf{X}_1 \mathbf{V}_3 - \mathbf{X}_2\|_F^2 + \|\mathbf{X}_3 \mathbf{V}_1^\top - \mathbf{X}_1 - \mathbf{X}_2 \mathbf{V}_3^\top\|_F^2.$$

Defining

$$\mathbf{H} = \begin{pmatrix} \mathbf{I} & \mathbf{V}_3 & -\mathbf{V}_1 \\ \mathbf{V}_3^\top & \mathbf{I} & -\mathbf{V}_2 \\ -\mathbf{V}_1^\top & -\mathbf{V}_2^\top & \mathbf{I} \end{pmatrix},$$

the optimization problem can be rewritten as

$$\min_{\mathbf{H} \in \mathbb{R}^{3D \times 3D}} \|(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)\mathbf{H}\|_F^2, \text{ subject to } \mathbf{H} = \mathbf{H}^\top, \mathbf{H}_{ii} = \mathbf{I}.$$

In general, for $K \geq 3$, this method can be formulated as

$$\min_{\mathbf{H} \in \mathbb{R}^{KD \times KD}} \text{tr}(\mathbf{C}\mathbf{H}^2), \text{ subject to } \mathbf{H} = \mathbf{H}^\top \text{ and } \mathbf{H}_{ii} = \mathbf{I} \text{ for all } 1 \leq i \leq K,$$

where \mathbf{H}_{ij} represents the ij -th $D \times D$ block of \mathbf{H} .

5. The LS+ approach with constraints on the operator norm of \mathbf{H}_{ij} (C-LS+):

$$\min_{\mathbf{H} \in \mathbb{R}^{KD \times KD}} \text{tr}(\mathbf{C}\mathbf{H}^2), \text{ subject to } \mathbf{H} = \mathbf{H}^\top, \mathbf{H}_{ii} = \mathbf{I} \text{ and } \|\mathbf{H}_{ij}\| \leq 1 \text{ for all } 1 \leq i, j \leq K.$$

To compare the SDP/LS+/C-LS+ approaches, we summarize their objective functions and their constraints in Table 1. There are two main differences. First, the objective functions are different. However, since $\text{tr}(\mathbf{C}\mathbf{H}) = 0$ if and only if $\text{tr}(\mathbf{C}\mathbf{H}^2) = 0$ (considering $\mathbf{C} \succcurlyeq \mathbf{0}$ and $\mathbf{H} \succcurlyeq \mathbf{0}$), this difference does not affect the property of exact recovery. Second, the constraints of the SDP approach are more restrictive than those of the C-LS+ approach ($\mathbf{H}_{ii} = \mathbf{H}_{jj} = \mathbf{I}$ and $\mathbf{H} \succcurlyeq \mathbf{0}$ imply $\|\mathbf{H}_{ij}\| \leq 1$), which is more restrictive than the C-LS approach. This observation partially justifies the fact that SDP performs better than C-LS+, and C-LS+ performs better than C-LS. However, these interpretations do not justify the empirical finding in Figures 1 and 3 that C-LS+ and SDP behave very similarly in the absence of noise. We leave the explanation of this observation as an open question.

Table 1: Comparison between SDP, LS+ and C-LS+ approaches.

	objective function	common constraint	other constraints
SDP	$\text{tr}(\mathbf{C}\mathbf{H})$	$\mathbf{H}_{ii} = \mathbf{I}, \mathbf{H} = \mathbf{H}^\top$	$\mathbf{H} \succcurlyeq \mathbf{0}$
LS+	$\text{tr}(\mathbf{C}\mathbf{H}^2)$		$\ \mathbf{H}_{ij}\ \leq 1$
C-LS+	$\text{tr}(\mathbf{C}\mathbf{H}^2)$		

Among these optimization approaches, the LS method has an explicit solution by decomposing it into D sub-problems, where each is a regression problem that estimates KD regression parameters. All other methods are convex and can be solved by CVX [15], where the default solver SeDuMi is used [16]. While the LS+ approach can also be written as a least squares problem with an explicit solution, this problem is not decomposable (unlike the LS method).

When the solution matrices of LS/C-LS are not orthogonal, they are rounded to the nearest orthogonal matrices using the approach in [14]. The rounding procedure of LS+/C-LS+ is the same as that of the SDP method.

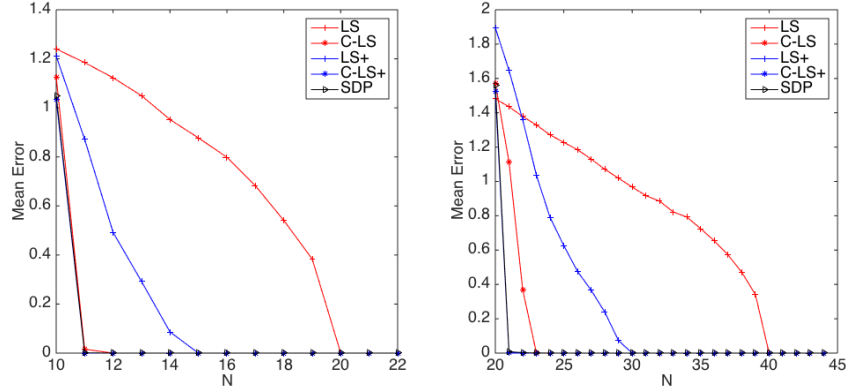


Figure 1: The dependence of the mean recovery error (over 50 runs) with respect to N , when $D = 10$ (left panel) and $D = 20$ (right panel). The y -axis represents the mean recovery error of \mathbf{V}_1 in Frobenius norm.

In the first simulation, we aim to find the size of N such that the orthogonal matrices be exactly recovered by the suggested algorithms for $K = 3$. We let $D = 10$ or 20 and choose various values for N , and record the mean recovery error of \mathbf{V}_1 (in Frobenius norm) over 50 repeated simulations in Figure 1. The performance of LS verifies our theoretical analysis: it recovers the orthogonal matrices for $N \geq 2D$. LS fails when $N < 2D$ because the null space of $[\mathbf{X}_1, \mathbf{X}_2]$ is nontrivial and there are infinite solutions. Besides, LS+ succeeds when $N \geq 3D/2$. SDP and C-LS+ are the best approaches and they succeed when $N \geq D + 1$, which verifies Theorem 2.2.

In the second simulation, we test the stability of the suggested algorithms when $K = 3$ and the measurement matrix \mathbf{X}_3 is contaminated elementwisely by Gaussian noise $\mathcal{N}(0, \sigma^2)$. We use the setting $N = 12, 16, 22$, $D = 10$ and $\sigma = 0.01$ or 0.1 and record the mean recovery error over 50 runs in Table 2, which shows that the SDP relaxation approach is more stable to noise than competing approaches. This motivates our interest in studying the SDP approach.

Table 2: The mean recovery error over 50 runs in the noisy setting for $K = 3$ and $D = 10$.

N	σ	SDP	C-LS+	LS+	C-LS	LS
12	0.01	0.071	0.076	0.482	0.508	2.260
16	0.01	0.026	0.031	0.037	0.059	1.926
22	0.01	0.018	0.021	0.020	0.030	0.077
12	0.1	0.742	0.742	1.088	0.880	2.341
16	0.1	0.261	0.328	0.399	0.459	2.034
22	0.1	0.175	0.217	0.216	0.262	0.834

In the third simulation, we compare these methods for $K = 5$ and $D = 5, 10$.

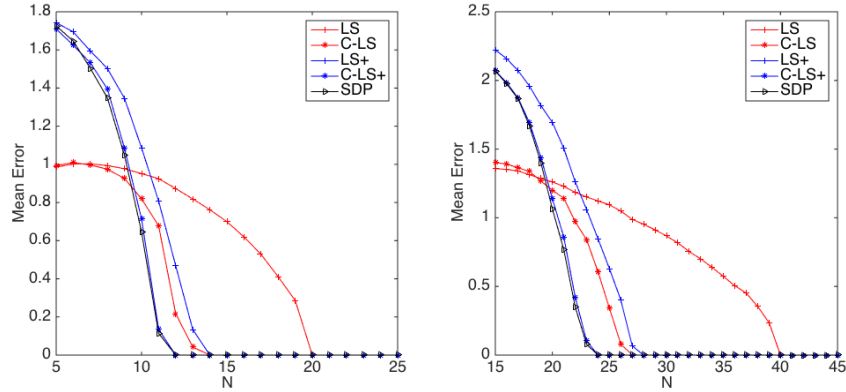


Figure 2: The dependence of the mean recovery error (over 50 runs) with respect to N , when $K = 5$, $D = 5$ (left panel) and $K = 5$, $D = 10$ (right panel). The y -axis represents the mean recovery error of \mathbf{V}_1 in Frobenius norm.

The results are shown in Figure 3. This simulation verifies Theorem 2.3 by showing that the SDP approach successfully recovers the orthogonal matrices for $N \geq (K - 2)D + 1$. Indeed, the empirical performance of the SDP approach is even better: it recovers $\{\mathbf{V}_i\}_{i=1}^5$ at $N = 12$ and 25 respectively, which are smaller than $(K - 2)D + 1$. Compared with LS/LS+/C-LS, the SDP and C-LS+ approaches recover the orthogonal matrices with smaller N .

At last, we record the running time for all approaches in Table 3. Although the running time is not the main focus of this paper, and CVX is not optimized for the approaches, this table gives a sense of the running times. Table 3 clearly shows that the LS approach is much faster than the other approaches, and the SDP approach is consistently faster than C-LS+. We suspect that it is due to the fact that SDP has fewer constraints, even though the constraint of SDP is more restrictive than that of C-LS+.

Table 3: The average running time (in seconds) when $\sigma = 0.1$, $K = 3$, $D = 10$, $N = 15$ (first row) and $\sigma = 0.1$, $K = 5$, $D = 10$, $N = 35$ (second row).

SDP	C-LS+	LS+	C-LS	LS
0.64	0.76	0.20	0.48	0.0005
5.33	6.58	0.62	0.93	0.0017

4. Proofs of main results

In this section, we first provide the proof for Theorem 2.2, assuming Theorem 2.1, and then provide the proofs for Theorems 2.1 and 2.3. The main reason for this organization is that, given Theorem 2.1 (whose proof is more technical), the proof of Theorem 2.2 is rather straightforward. This organization would

also emphasize the importance of Theorem 2.1, which plays an important role in the proof of Theorems 2.2.

4.1. Proof of Theorem 2.2

Part (a) follows from the result in part (b) with $\epsilon = 0$, so it is sufficient to prove part (b).

In the proof of part (b), we first claim that it is sufficient to prove the case $\mathbf{V}_1 = \mathbf{V}_2 = -\mathbf{I}$, i.e., when $\mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3 = \mathbf{0}$ and $\|\hat{\mathbf{X}}_i - \mathbf{X}_i\|_F \leq \epsilon$ for $1 \leq i \leq 3$, then the SDP method recovers $\hat{\mathbf{V}}_1$ and $\hat{\mathbf{V}}_2$ such that $\|\hat{\mathbf{V}}_i + \mathbf{I}\|_F \leq C\sqrt{\epsilon}$ for $i = 1, 2$.

This result implies that if the input of the SDP method is $(-\hat{\mathbf{X}}_1 \mathbf{V}_1, -\hat{\mathbf{X}}_2 \mathbf{V}_2, \hat{\mathbf{X}}_3)$ and the output is denoted by $(\tilde{\mathbf{V}}_1, \tilde{\mathbf{V}}_2)$, then $\|\tilde{\mathbf{V}}_i + \mathbf{I}\|_F \leq C\sqrt{\epsilon}$ for $i = 1, 2$.

Additionally, it can be verified that if the SDP method outputs $(\mathbf{V}_1, \mathbf{V}_2)$ when the input is $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$, \mathbf{U}_1 and \mathbf{U}_2 are two orthogonal matrices, then the output for $(\mathbf{X}_1 \mathbf{U}_1, \mathbf{X}_2 \mathbf{U}_2, \mathbf{X}_3)$ would be $(\mathbf{U}_1^\top \mathbf{V}_1, \mathbf{U}_2^\top \mathbf{V}_2)$. Applying this observation, we have $\hat{\mathbf{V}}_i = -\hat{\mathbf{V}}_i \mathbf{V}_i^\top$. Combining it with $\|\hat{\mathbf{V}}_i + \mathbf{I}\|_F \leq C\sqrt{\epsilon}$ for $i = 1, 2$, the theorem is proved.

The rest of the proof will assume $\mathbf{V}_1 = \mathbf{V}_2 = -\mathbf{I}$ and $\mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3 = \mathbf{0}$. We represent the noisy setting in (8) by $\hat{\mathbf{X}}, \hat{\mathbf{C}}$ and $\hat{\mathbf{H}}$, the clean setting by \mathbf{X}, \mathbf{C} and \mathbf{H} , and write the decomposition of \mathbf{H} and $\hat{\mathbf{H}}$ by $\mathbf{H} = \mathbf{U}\mathbf{U}^\top$ and $\hat{\mathbf{H}} = \hat{\mathbf{U}}\hat{\mathbf{U}}^\top$.

Since $\hat{\mathbf{U}}_i \hat{\mathbf{U}}_i^\top = \mathbf{I}$, $\|\hat{\mathbf{U}}\| \leq \sum_{i=1}^3 \|\hat{\mathbf{U}}_i\| = 3$, (30) implies

$$\|\mathbf{X}\hat{\mathbf{U}}\|_F - \|\hat{\mathbf{X}}\hat{\mathbf{U}}\|_F \leq \|(\mathbf{X} - \hat{\mathbf{X}})\hat{\mathbf{U}}\|_F \leq 3\|\mathbf{X} - \hat{\mathbf{X}}\|_F \leq 9\epsilon \quad (12)$$

and following the same argument,

$$\|\hat{\mathbf{X}}\hat{\mathbf{U}}\|_F \leq 9\epsilon + \|\mathbf{X}\hat{\mathbf{U}}\|_F. \quad (13)$$

Since $\hat{\mathbf{H}} = \hat{\mathbf{U}}\hat{\mathbf{U}}^\top$ is the minimizer of the SDP problem, we have $\|\hat{\mathbf{X}}\hat{\mathbf{U}}\|_F \leq \|\mathbf{X}\hat{\mathbf{U}}\|_F$. Combining it with $\|\mathbf{X}\hat{\mathbf{U}}\|_F = 0$ and (13),

$$\|\hat{\mathbf{X}}\hat{\mathbf{U}}\|_F \leq 9\epsilon. \quad (14)$$

Combining (12), (14) with Theorem 2.1,

$$9\epsilon \geq \|\hat{\mathbf{X}}\hat{\mathbf{U}}\|_F \geq \|\mathbf{X}\hat{\mathbf{U}}\|_F - 9\epsilon \geq c(\mathbf{X})\|\hat{\mathbf{U}}^\top \mathbf{P}_{\mathbf{L}_\perp^\perp}\|_F^2 - 9\epsilon, \quad (15)$$

so we have

$$\|\hat{\mathbf{U}}^\top \mathbf{P}_{\mathbf{L}_\perp^\perp}\|_F \leq \sqrt{\frac{18\epsilon}{c(\mathbf{X})}}. \quad (16)$$

Combining it with Lemma 4.2,

$$\begin{aligned} \|\hat{\mathbf{U}}_1 \hat{\mathbf{U}}_3^\top - \mathbf{I}\|_F &= \|\hat{\mathbf{U}}_1 - \hat{\mathbf{U}}_3\|_F = \|\hat{\mathbf{U}}^\top (\mathbf{I}, \mathbf{0}, -\mathbf{I})\|_F = \|\hat{\mathbf{U}}^\top \mathbf{P}_{\mathbf{L}_\perp^\perp} \mathbf{P}_{\mathbf{L}_\perp^\perp}^\top (\mathbf{I}, \mathbf{0}, -\mathbf{I})\|_F \\ &\leq \|\hat{\mathbf{U}}^\top \mathbf{P}_{\mathbf{L}_\perp^\perp}\|_F \|\mathbf{P}_{\mathbf{L}_\perp^\perp}^\top (\mathbf{I}, \mathbf{0}, -\mathbf{I})\| \leq 2\sqrt{\frac{18\epsilon}{c(\mathbf{X})}}. \end{aligned}$$

Since the post-processing step $f(\mathbf{Z})$ is a continuous and differentiable function with respect to \mathbf{Z} and $f(-\mathbf{I}) = -\mathbf{I}$, the difference between the recovered orthogonal matrix $\hat{\mathbf{V}}_1$ and $\mathbf{V}_1 = -\mathbf{I}$, $\|\hat{\mathbf{V}}_1 + \mathbf{I}\|_F$, is bounded above by $C\sqrt{\epsilon}$ when $c(\mathbf{X}) > 0$. Similarly we have $\|\hat{\mathbf{V}}_2 + \mathbf{I}\|_F < C\sqrt{\epsilon}$.

4.2. Proof of Theorem 2.1

Proof. The main idea of the proof is to investigate $\tilde{\mathbf{U}}^*$, which is defined to be nearest matrix to \mathbf{U} in the set $\tilde{\mathcal{U}}$ defined in (17). Then we represent $\tilde{\mathbf{U}}^*$ in the form of (21), and show that $\|\mathbf{L}\|_F$ is bounded by $C_1\|\mathbf{X}\mathbf{U}\|_F$, for some $C_1 > 0$. With additional bounds $\|\mathbf{K}\mathbf{Y}_1\|_F, \|\mathbf{K}\mathbf{Y}_2\|_F \leq C_2\sqrt{\|\mathbf{X}\mathbf{U}\|_F}$ and

$$\|\tilde{\mathbf{U}}^{*\top} \mathbf{P}_{\mathbf{L}_1^\perp}\|_F \leq \left\| \begin{pmatrix} (\mathbf{L}\mathbf{Y}_1)^\top & (\mathbf{L}\mathbf{Y}_2)^\top & \mathbf{0} \\ (\mathbf{K}\mathbf{Y}_1)^\top & (\mathbf{K}\mathbf{Y}_2)^\top & \mathbf{0} \end{pmatrix} \right\|_F,$$

we will show that $\|\tilde{\mathbf{U}}^{*\top} \mathbf{P}_{\mathbf{L}_1^\perp}\|_F$ is bounded above by a function of $\|\mathbf{X}\mathbf{U}\|_F$ in (24). By analyzing the properties of $\tilde{\mathbf{U}}^*$, the same statement holds for $\|\mathbf{U}\mathbf{P}_{\mathbf{L}_1^\perp}\|_F$, and the theorem is proved.

We first remark that it is sufficient to prove the case $N = D + 1$. If this is true, then for $N > D + 1$, (9) holds when \mathbf{X} is replaced by \mathbf{X}' , the submatrix consists of the first $D + 1$ rows of \mathbf{X} . Since $\|\mathbf{X}\mathbf{U}\|_F \geq \|\mathbf{X}'\mathbf{U}\|_F$, (9) is proved. Therefore, for the rest of the proof, we assume $N = D + 1$.

We first define the following:

$$\begin{aligned} \bar{\mathcal{U}} &= \{\bar{\mathbf{U}} \in \mathbb{R}^{3D \times k} : \mathbf{X}\bar{\mathbf{U}} = \mathbf{0}\}, \\ \tilde{\mathcal{U}} &= \{\tilde{\mathbf{U}} \in \mathbb{R}^{3D \times k} : \mathbf{X}\tilde{\mathbf{U}} = \mathbf{0}, \tilde{\mathbf{U}}_3 = \mathbf{U}_3\}, \end{aligned} \quad (17)$$

and in (17), $\tilde{\mathbf{U}}_3$ and \mathbf{U}_3 represent the submatrix consists of the last D rows of $\tilde{\mathbf{U}}$ and \mathbf{U} . We also define the distances between two matrices and the distances between a matrix and a set by

$$\text{dist}(\mathbf{U}, \mathbf{U}') = \|\mathbf{U} - \mathbf{U}'\|_F, \quad \text{dist}(\mathbf{U}, \mathcal{U}) = \min_{\mathbf{U}' \in \mathcal{U}} \|\mathbf{U} - \mathbf{U}'\|_F.$$

Then we have

$$\text{dist}(\mathbf{U}, \bar{\mathcal{U}}) \geq C \text{dist}(\mathbf{U}, \tilde{\mathcal{U}}), \quad (18)$$

for $C = 1/2$. The proof of (18) is deferred to Section 4.2.1.

Assuming that $\sigma_{\min}(\mathbf{X})$ is the smallest singular value of \mathbf{X} , and for any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\text{Sp}(\mathbf{A})$ represents the subspace spanned by the row vectors of \mathbf{A} in \mathbb{R}^n , then we have

$$\|\mathbf{X}\mathbf{U}\|_F \geq \sigma_{\min}(\mathbf{X}) \|\mathbf{P}_{\text{Sp}(\mathbf{X})}^\top \mathbf{U}\|_F = \sigma_{\min}(\mathbf{X}) \text{dist}(\mathbf{U}, \bar{\mathcal{U}}) \geq C \sigma_{\min}(\mathbf{X}) \text{dist}(\mathbf{U}, \tilde{\mathcal{U}}), \quad (19)$$

where the last inequality follows from (18).

Assuming that $\tilde{\mathbf{U}}^* = \arg \min_{\tilde{\mathbf{U}} \in \tilde{\mathcal{U}}} \text{dist}(\tilde{\mathbf{U}}, \mathbf{U})$, then using $\begin{pmatrix} \mathbf{U}_3 \\ \mathbf{U}_3 \\ \mathbf{U}_3 \end{pmatrix} \in \tilde{\mathcal{U}}$ we have

$$\text{dist}(\tilde{\mathbf{U}}^*, \mathbf{U}) \leq \text{dist}\left(\begin{pmatrix} \mathbf{U}_3 \\ \mathbf{U}_3 \\ \mathbf{U}_3 \end{pmatrix}, \mathbf{U}\right) \leq \sqrt{8D}. \quad (20)$$

Now let us investigate $\text{dist}(\tilde{\mathbf{U}}^*, \mathcal{U})$ further. If \mathbf{Y}_1 and $\mathbf{Y}_2 \in \mathbb{R}^{(2D-N) \times D}$ are chosen such that $[\mathbf{X}_1, \mathbf{X}_2][\mathbf{Y}_1, \mathbf{Y}_2]^\top = \mathbf{0}$, then using $\mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3 = \mathbf{0}$, there exist $\mathbf{L} \in \mathbb{R}^{D \times (3D-N)}$, $\mathbf{K} \in \mathbb{R}^{(k-D) \times (3D-N)}$ and an orthogonal matrix $\mathbf{U}'_3 \in \mathbb{R}^{k \times k}$ such that

$$\tilde{\mathbf{U}}^* = \begin{pmatrix} \mathbf{I} + \mathbf{L}\mathbf{Y}_1 & \mathbf{K}\mathbf{Y}_1 \\ \mathbf{I} + \mathbf{L}\mathbf{Y}_2 & \mathbf{K}\mathbf{Y}_2 \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \mathbf{U}'_3. \quad (21)$$

That is, if we write $\tilde{\mathbf{U}}^* = \begin{pmatrix} \tilde{\mathbf{U}}_1^* \\ \tilde{\mathbf{U}}_2^* \\ \tilde{\mathbf{U}}_3^* \end{pmatrix}$, then

$$\tilde{\mathbf{U}}_1^* = (\mathbf{I} + \mathbf{L}\mathbf{Y}_1, \mathbf{K}\mathbf{Y}_1) \mathbf{U}'_3, \quad \tilde{\mathbf{U}}_2^* = (\mathbf{I} + \mathbf{L}\mathbf{Y}_2, \mathbf{K}\mathbf{Y}_2) \mathbf{U}'_3, \quad \tilde{\mathbf{U}}_3^* = \mathbf{U}_3$$

($\tilde{\mathbf{U}}_3^* = \mathbf{U}_3$ follows from definition of $\tilde{\mathcal{U}}$). Since for any $\mathbf{U}_1 \in \mathbb{R}^{k \times D}$ with SVD decomposition $\mathbf{U}_1 = \mathbf{U}_{U_1} \boldsymbol{\Sigma}_{U_1} \mathbf{V}_{U_1}^\top$, the closest orthogonal matrix in $\mathbb{R}^{k \times D}$ is given by $\mathbf{U}_{U_1} \mathbf{V}_{U_1}^\top$, so the distance between $\tilde{\mathbf{U}}^*$ and \mathcal{U} is

$$\text{dist}(\tilde{\mathbf{U}}^*, \mathcal{U}) = \sqrt{\sum_{i=1}^3 \|\boldsymbol{\Sigma}_{\tilde{\mathbf{U}}_i^*} - \mathbf{I}\|_F^2}.$$

Applying (20), all singular values of $\tilde{\mathbf{U}}_1$, $\tilde{\mathbf{U}}_2$ and $\tilde{\mathbf{U}}_3$ (i.e., all diagonal entries of $\{\boldsymbol{\Sigma}_{\tilde{\mathbf{U}}_i^*}\}_{i=1}^3$) are smaller than $\sqrt{8D} + 1$. Let $C' = \sqrt{8D} + 2$, then $\text{dist}(\mathbf{U}, \mathcal{U})$ can be controlled as follows:

$$\begin{aligned} C'^2 \text{dist}^2(\mathbf{U}, \tilde{\mathcal{U}}) &= C'^2 \text{dist}^2(\tilde{\mathbf{U}}^*, \mathbf{U}) \geq C'^2 \text{dist}^2(\tilde{\mathbf{U}}^*, \mathcal{U}) = C'^2 \sum_{i=1}^3 \|\boldsymbol{\Sigma}_{\tilde{\mathbf{U}}_i^*} - \mathbf{I}\|_F^2 \\ &\geq \sum_{i=1}^3 \|\boldsymbol{\Sigma}_{\tilde{\mathbf{U}}_i^*}^2 - \mathbf{I}\|_F^2 = \|\tilde{\mathbf{U}}_1^* \tilde{\mathbf{U}}_1^{*\top} - \mathbf{I}\|_F^2 + \|\tilde{\mathbf{U}}_2^* \tilde{\mathbf{U}}_2^{*\top} - \mathbf{I}\|_F^2 + \|\tilde{\mathbf{U}}_3^* \tilde{\mathbf{U}}_3^{*\top} - \mathbf{I}\|_F^2 \\ &= \|\mathbf{L}\mathbf{Y}_1 + \mathbf{Y}_1^\top \mathbf{L}^\top - \mathbf{Y}_1^\top (\mathbf{L}^\top \mathbf{L} + \mathbf{K}^\top \mathbf{K}) \mathbf{Y}_1\|_F^2 \\ &\quad + \|\mathbf{L}\mathbf{Y}_2 + \mathbf{Y}_2^\top \mathbf{L}^\top - \mathbf{Y}_2^\top (\mathbf{L}^\top \mathbf{L} + \mathbf{K}^\top \mathbf{K}) \mathbf{Y}_2\|_F^2 \end{aligned} \quad (22)$$

$$\geq (\min(0, \lambda_{\min}(\mathbf{L}\mathbf{Y}_1 + \mathbf{Y}_1^\top \mathbf{L}^\top)))^2 + (\min(0, \lambda_{\min}(\mathbf{L}\mathbf{Y}_2 + \mathbf{Y}_2^\top \mathbf{L}^\top)))^2, \quad (23)$$

where the last inequality follows from the observation that $\mathbf{Y}_1^\top (\mathbf{L}^\top \mathbf{L} + \mathbf{K}^\top \mathbf{K}) \mathbf{Y}_1$ and $\mathbf{Y}_2^\top (\mathbf{L}^\top \mathbf{L} + \mathbf{K}^\top \mathbf{K}) \mathbf{Y}_2$ are positive semidefinite, and for any symmetric

matrix \mathbf{X} , the distance to the nearest positive semidefinite matrix (in Forbenius norm) is at least $-\min(0, \lambda_{\min}(\mathbf{X}))$.

Then (19) and (23) imply

$$\lambda_{\min}(\mathbf{L}\mathbf{Y}_1 + \mathbf{Y}_1^\top \mathbf{L}^\top) \geq -\frac{C'\|\mathbf{X}\mathbf{U}\|_F}{C\sigma_{\min}(\mathbf{X})}, \quad \lambda_{\min}(\mathbf{L}\mathbf{Y}_2 + \mathbf{Y}_2^\top \mathbf{L}^\top) \geq -\frac{C'\|\mathbf{X}\mathbf{U}\|_F}{C\sigma_{\min}(\mathbf{X})}.$$

Now we introduce an important lemma.

Lemma 4.1. *For $D > M$, $\mathbf{L} \in \mathbb{R}^{D \times M}$ and $\mathbf{Y}, \mathbf{Z} \in \mathbb{R}^{M \times D}$, if $\lambda_{\min}(\mathbf{L}\mathbf{Y} + \mathbf{Y}^\top \mathbf{L}^\top) \geq -\epsilon$ and $\lambda_{\min}(\mathbf{L}\mathbf{Z} + \mathbf{Z}^\top \mathbf{L}^\top) \geq -\epsilon$, where λ_{\min} represents the smallest eigenvalue, then $\|\mathbf{L}\|_F \leq \epsilon c(\mathbf{Y}, \mathbf{Z})$, where $c(\mathbf{Y}, \mathbf{Z}) > 0$ for generic \mathbf{Y} and \mathbf{Z} .*

The proof of Lemma 4.1 is rather technical and deferred to Section 4.4. Applying Lemma 4.1, for generic \mathbf{Y}_1 and \mathbf{Y}_2 (and as a result, for generic \mathbf{X}_1 and \mathbf{X}_2) there exists C_1 depending on \mathbf{Y}_1 and \mathbf{Y}_2 such that $\|\mathbf{L}\|_F \leq C_1\|\mathbf{X}\mathbf{U}\|_F$. In addition, we have

$$\begin{aligned} & \|\mathbf{L}\mathbf{Y}_1 + \mathbf{Y}_1^\top \mathbf{L}^\top - \mathbf{Y}_1^\top (\mathbf{L}^\top \mathbf{L} + \mathbf{K}^\top \mathbf{K}) \mathbf{Y}_1\|_F \\ & \geq \|\mathbf{Y}_1^\top \mathbf{K}^\top \mathbf{K} \mathbf{Y}_1\|_F - \|\mathbf{Y}_1^\top \mathbf{L}^\top \mathbf{L} \mathbf{Y}_1\|_F - \|\mathbf{L}\mathbf{Y}_1 + \mathbf{Y}_1^\top \mathbf{L}^\top\|_F, \end{aligned}$$

and $\|\mathbf{Y}_1^\top \mathbf{K}^\top \mathbf{K} \mathbf{Y}_1\|_F \geq \frac{1}{\sqrt{D}} \text{tr}(\mathbf{Y}_1^\top \mathbf{K}^\top \mathbf{K} \mathbf{Y}_1) = \frac{1}{\sqrt{D}} \|\mathbf{K} \mathbf{Y}_1\|_F^2$. As a result, there exists C_2 depending on $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{X}$ such that if $\|\mathbf{K} \mathbf{Y}_1\|_F > C_2 \sqrt{\|\mathbf{X}\mathbf{U}\|_F}$ then

$$\|\mathbf{L}\mathbf{Y}_1 + \mathbf{Y}_1^\top \mathbf{L}^\top - \mathbf{Y}_1^\top (\mathbf{L}^\top \mathbf{L} + \mathbf{K}^\top \mathbf{K}) \mathbf{Y}_1\|_F \geq \frac{C'\|\mathbf{X}\mathbf{U}\|_F}{C\sigma_{\min}(\mathbf{X})},$$

which violates (22) and (19). Therefore, by contradiction we proved $\|\mathbf{K} \mathbf{Y}_1\|_F \leq C_2 \sqrt{\|\mathbf{X}\mathbf{U}\|_F}$, and similarly, $\|\mathbf{K} \mathbf{Y}_2\|_F \leq C_2 \sqrt{\|\mathbf{X}\mathbf{U}\|_F}$. Combining it with $\|\mathbf{L}\|_F \leq C_1\|\mathbf{X}\mathbf{U}\|_F$, we have

$$\begin{aligned} \|\tilde{\mathbf{U}}^{*\top} \mathbf{P}_{\mathbf{L}_\perp^\perp}\|_F &= \|\tilde{\mathbf{U}}^{*\top} \mathbf{P}_{\mathbf{L}_\perp^\perp}\|_F = \left\| \begin{pmatrix} (\mathbf{L}\mathbf{Y}_1)^\top & (\mathbf{L}\mathbf{Y}_2)^\top & \mathbf{0} \\ (\mathbf{K}\mathbf{Y}_1)^\top & (\mathbf{K}\mathbf{Y}_2)^\top & \mathbf{0} \end{pmatrix} \mathbf{P}_{\mathbf{L}_\perp^\perp} \right\|_F \\ &\leq \left\| \begin{pmatrix} (\mathbf{L}\mathbf{Y}_1)^\top & (\mathbf{L}\mathbf{Y}_2)^\top & \mathbf{0} \\ (\mathbf{K}\mathbf{Y}_1)^\top & (\mathbf{K}\mathbf{Y}_2)^\top & \mathbf{0} \end{pmatrix} \right\|_F \leq \|\mathbf{L}\mathbf{Y}_1\|_F + \|\mathbf{L}\mathbf{Y}_2\|_F + \|\mathbf{K}\mathbf{Y}_1\|_F + \|\mathbf{K}\mathbf{Y}_2\|_F \\ &\leq \|\mathbf{L}\|_F (\|\mathbf{Y}_1\| + \|\mathbf{Y}_2\|) + \|\mathbf{K}\mathbf{Y}_1\|_F + \|\mathbf{K}\mathbf{Y}_2\|_F \\ &\leq (\|\mathbf{Y}_1\| + \|\mathbf{Y}_2\|) C_1 \|\mathbf{X}\mathbf{U}\|_F + 2C_2 \sqrt{\|\mathbf{X}\mathbf{U}\|_F}. \end{aligned} \tag{24}$$

Recall that C_1 and C_2 only depends on \mathbf{X}_1 and \mathbf{X}_2 (\mathbf{X} , \mathbf{Y}_1 , and \mathbf{Y}_2 are generated from \mathbf{X}_1 and \mathbf{X}_2), combining (19) and (24) we have

$$\begin{aligned} \|\mathbf{U}^\top \mathbf{P}_{\mathbf{L}_\perp^\perp}\|_F &\leq \|\tilde{\mathbf{U}}^{*\top} \mathbf{P}_{\mathbf{L}_\perp^\perp}\|_F + \|(\mathbf{U} - \tilde{\mathbf{U}}^*)^\top \mathbf{P}_{\mathbf{L}_\perp^\perp}\|_F \leq \|\tilde{\mathbf{U}}^{*\top} \mathbf{P}_{\mathbf{L}_\perp^\perp}\|_F + \|\mathbf{U} - \tilde{\mathbf{U}}^*\|_F \\ &\leq (\|\mathbf{Y}_1\| + \|\mathbf{Y}_2\|) C_1 \|\mathbf{X}\mathbf{U}\|_F + 2C_2 \sqrt{\|\mathbf{X}\mathbf{U}\|_F} + \frac{1}{C\sigma_{\min}(\mathbf{X})} \|\mathbf{X}\mathbf{U}\|_F. \end{aligned}$$

Considering that $\|\mathbf{U}\| \leq 3$ and $\|\mathbf{X}\mathbf{U}\|_F \leq \|\mathbf{U}\| \|\mathbf{X}\|_F \leq 3\|\mathbf{X}\|_F$, so if we let $C_4 = 3\|\mathbf{X}\|_F$, then

$$\|\mathbf{U}^\top \mathbf{P}_{\mathbf{L}_\perp^\perp}\|_F \leq \left((\|\mathbf{Y}_1\| + \|\mathbf{Y}_2\|) C_1 \sqrt{C_4} + 2C_2 + \frac{1}{C\sigma_{\min}(\mathbf{X})} \sqrt{C_4} \right) \sqrt{\|\mathbf{X}\mathbf{U}\|_F},$$

and Theorem 2.1 is proved. \square

4.2.1. Proof of (18)

Suppose that $\bar{\mathbf{U}}^* = \arg \min_{\bar{\mathbf{U}} \in \bar{\mathcal{U}}} \text{dist}(\bar{\mathbf{U}}, \mathbf{U})$ and

$$\mathbf{U} - \bar{\mathbf{U}}^* = \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \mathbf{V}_3 \end{pmatrix}, \text{ where } \mathbf{V}_i \in \mathbb{R}^{D \times k} \text{ for } i = 1, 2, 3,$$

then

$$\tilde{\mathbf{U}} = \bar{\mathbf{U}}^* + \begin{pmatrix} \mathbf{V}_3 \\ \mathbf{V}_3 \\ \mathbf{V}_3 \end{pmatrix}$$

would satisfies that $\tilde{\mathbf{U}} \in \tilde{\mathcal{U}}$, and as a result,

$$\begin{aligned} \text{dist}(\mathbf{U}, \bar{\mathcal{U}}) &= \text{dist}(\mathbf{U}, \bar{\mathbf{U}}^*) = \sqrt{\|\mathbf{V}_1\|_F^2 + \|\mathbf{V}_2\|_F^2 + \|\mathbf{V}_3\|_F^2} \geq C \sqrt{\|\mathbf{V}_1 + \mathbf{V}_3\|_F^2 + \|\mathbf{V}_2 + \mathbf{V}_3\|_F^2} \\ &= C \text{dist}(\mathbf{U}, \tilde{\mathbf{U}}) \geq C \text{dist}(\mathbf{U}, \tilde{\mathcal{U}}), \end{aligned}$$

where C can chosen to be $1/2$.

4.3. Proof of Theorem 2.3

Proof. We start with the same argument as in the proof of Theorem 2.2 and assume $\mathbf{V}_i = -\mathbf{I}$ for all $1 \leq i \leq K-1$ and $\sum_{i=1}^K \mathbf{X}_i = \mathbf{0}$. Then the proof can be divided into three steps. First, we show that it is sufficient to prove that a property defined in (26) is satisfied for generic $\mathbf{Y} \in \mathbb{R}^{p \times KD}$ where $p = \max((K-1)D - N, 0)$. Then we establish that (26) indeed holds for generic \mathbf{Y} . To this end, secondly, we show that any \mathbf{Y} that does not satisfy this property lies in a certain set. Finally, in the third step, we show that this set is of measure zero.

4.3.1. Step 1: reduction of the problem to the property (26)

By the assumption that $\mathbf{V}_i = -\mathbf{I}$ for all $1 \leq i \leq K-1$, $\mathbf{X}_1 + \dots + \mathbf{X}_K = \mathbf{0}$, and

$$\text{tr}(\mathbf{H}(\mathbf{I}, \dots, \mathbf{I})^\top (\mathbf{I}, \dots, \mathbf{I})) = \text{tr}((\mathbf{I}, \dots, \mathbf{I}) \mathbf{H} (\mathbf{I}, \dots, \mathbf{I})^\top) = \|\mathbf{X}_1 + \dots + \mathbf{X}_K\|_F^2 = 0.$$

Considering that $\text{tr}(\mathbf{H}\mathbf{C}) \geq 0$ for any $\mathbf{H} \succcurlyeq \mathbf{0}$, if the solution to the SDP problem is not uniquely given by $(\mathbf{I}, \dots, \mathbf{I})^\top (\mathbf{I}, \dots, \mathbf{I})$, then there exists $\mathbf{H} \neq (\mathbf{I}, \dots, \mathbf{I})^\top (\mathbf{I}, \dots, \mathbf{I})$ such that $\text{tr}(\mathbf{C}\mathbf{H}) = 0$. Let $\mathbf{H} = \mathbf{U}\mathbf{U}^\top$ for a matrix $\mathbf{U} \in \mathbb{R}^{KD \times k}$, then using the properties of \mathbf{H} , we have $\mathbf{U} \in \mathcal{U}$ for

$$\mathcal{U} = \left\{ \mathbf{U} = \begin{pmatrix} \mathbf{U}_1 \\ \vdots \\ \mathbf{U}_K \end{pmatrix} : \mathbf{U}_i \in \mathbb{R}^{D \times k}, \mathbf{U}_i \mathbf{U}_i^\top = \mathbf{I}, \forall 1 \leq i \leq K, \mathbf{P}_{\mathbf{L}_1^\perp}^\top \mathbf{U} \neq \mathbf{0} \right\},$$

L_1 defined by

$$L_1 = \{\mathbf{z} \in \mathbb{R}^{KD} : \mathbf{z} = (\mathbf{x}, \mathbf{x}, \dots, \mathbf{x}) \text{ for some } \mathbf{x} \in \mathbb{R}^D\},$$

and $\mathbf{P}_{L_1^\perp}^\top \mathbf{U} \neq \mathbf{0}$ means that $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_K$ are not all the same.

Since $\text{tr}(\mathbf{C}\mathbf{H}) = 0$, we have $\|\mathbf{X}\mathbf{U}\|_F = 0$. Let $\text{Sp}(\mathbf{A})$ and $\text{Col}(\mathbf{A})$ be the subspaces spanned by the row vectors of \mathbf{A} and the column vectors of \mathbf{A} respectively, then

$$\text{Sp}(\mathbf{X})^\perp \supseteq \text{Col}(\mathbf{U}). \quad (25)$$

For any two subspace L and L' , we use $L + L'$ to represent the subspace $\{\mathbf{x} + \mathbf{y} : \mathbf{x} \in L, \mathbf{y} \in L'\}$, then we claim that to prove Theorem 2.3, it is sufficient to prove the following statement with $p = \max((K-1)D - N, 0)$:

$$\text{For generic } \mathbf{Y} \in \mathbb{R}^{p \times KD}, \text{Col}(\mathbf{U}) \not\subseteq \text{Sp}(\mathbf{Y}) + L_1 \text{ for every } \mathbf{U} \in \mathcal{U}. \quad (26)$$

The argument is as follows. Since $\sum_{i=1}^K \mathbf{X}_i = \mathbf{0}$,

$$\begin{aligned} \dim(\text{Sp}(\mathbf{X})) &= \text{rank}(\mathbf{X}) = \text{rank}([\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{K-1}, \mathbf{X}_K]) \\ &= \text{rank}([\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{K-1}]), \end{aligned}$$

which is $\min((K-1)D, N)$ for generic $\{\mathbf{X}_i\}_{i=1}^{K-1} \in \mathbb{R}^{N \times D}$, so $\text{Sp}(\mathbf{X})$ is a generic $\min((K-1)D, N)$ -dimensional subspace in L_1^\perp (and L_1^\perp is a subspace of \mathbb{R}^{DK}).

So, $\text{Sp}(\mathbf{X})^\perp$ is the sum of L_1 and a generic p -dimensional subspace in \mathbb{R}^{KD} . Therefore, for generic $\{\mathbf{X}\}_{i=1}^{K-1}$, $\text{Sp}(\mathbf{X})^\perp$ is equivalent to $\text{Sp}(\mathbf{Y}) + L_1$, where \mathbf{Y} is a generic matrix of size $p \times KD$. If (26) holds, then (25) would not hold for generic $\{\mathbf{X}\}_{i=1}^{K-1}$ and every $\mathbf{U} \in \mathcal{U}$. By the analysis before (25), the solution to the SDP problem is uniquely given by $(\mathbf{I}, \dots, \mathbf{I})^\top (\mathbf{I}, \dots, \mathbf{I})$, and Theorem 2.3 is proved.

4.3.2. Step 2: finding matrices that do not satisfy (26)

In this part we show that every \mathbf{Y} violating (26) lies in the set $\cup_{d=1}^p \mathcal{M}_d$, where \mathcal{M}_d is the range of the function

$$g_d(\{\mathbf{Y}_i\}_{i=0}^{K-1}, \mathbf{Z}, L, L_0) = \mathbf{P}_L [(\mathbf{Y}_1 \mathbf{P}_{L_0}^\top, \dots, \mathbf{Y}_K \mathbf{P}_{L_0}^\top) + (\mathbf{Y}_0, \dots, \mathbf{Y}_0)] + \mathbf{P}_{L^\perp} \mathbf{Z}.$$

The domain of the function g_d is as follows: $\mathbf{Y}_i \in \mathbb{R}^{d \times d}$ for $1 \leq i \leq K-1$, $\mathbf{Y}_0 \in \mathbb{R}^{d \times D}$, $\mathbf{Z} \in \mathbb{R}^{(p-d) \times KD}$, L_0 is a d -dimensional subspace in \mathbb{R}^D , and L is a d -dimensional subspace in \mathbb{R}^p . In addition, $\mathbf{Y}_K = -\sum_{i=1}^{K-1} \mathbf{Y}_i$.

For every \mathbf{Y} violating (26), there exists $\mathbf{U} \in \mathcal{U}$ such that $\text{Col}(\mathbf{U}) \subseteq \text{Sp}(\mathbf{Y}) + L_1$. We let $\mathbf{U} = \mathbf{U}^{(1)} + \mathbf{U}^{(2)}$, where the columns of $\mathbf{U}^{(1)}$ lie in L_1 and the columns of $\mathbf{U}^{(2)}$ are orthogonal to L_1 . As a result, $\text{Col}(\mathbf{U}^{(2)})$ is a subspace in \mathbb{R}^{KD} that intersects L_1 only at origin, and we let $d = \dim(\text{Col}(\mathbf{U}^{(2)}))$. We have $d \geq 1$, since otherwise $\mathbf{U}^{(2)}$ is a zero matrix, and $\mathbf{P}_{L_1^\perp}^\top \mathbf{U} = \mathbf{P}_{L_1^\perp}^\top \mathbf{U}^{(1)} = 0$, which contradicts $\mathbf{U} \in \mathcal{U}$.

Denote

$$\mathbf{U}^{(1)} = \begin{pmatrix} \mathbf{U}_1^{(1)} \\ \vdots \\ \mathbf{U}_K^{(1)} \end{pmatrix} \quad \text{and} \quad \mathbf{U}^{(2)} = \begin{pmatrix} \mathbf{U}_1^{(2)} \\ \vdots \\ \mathbf{U}_K^{(2)} \end{pmatrix},$$

then $\mathbf{U}_1^{(1)} = \mathbf{U}_2^{(1)} = \dots = \mathbf{U}_K^{(1)}$. Recall for all $1 \leq i \leq K$, $\mathbf{U}_i^{(1)} \mathbf{U}_i^{(1)\top} + \mathbf{U}_i^{(2)} \mathbf{U}_i^{(2)\top} = \mathbf{U}_i \mathbf{U}_i^\top = \mathbf{I}$, we have

$$\mathbf{U}_1^{(2)} \mathbf{U}_1^{(2)\top} = \mathbf{U}_2^{(2)} \mathbf{U}_2^{(2)\top} = \dots = \mathbf{U}_K^{(2)} \mathbf{U}_K^{(2)\top}.$$

Therefore,

$$\text{Col}(\mathbf{U}_1^{(2)}) = \text{Col}(\mathbf{U}_2^{(2)}) = \dots = \text{Col}(\mathbf{U}_K^{(2)}), \quad (27)$$

and

$$\dim(\text{Col}(\mathbf{U}_1^{(2)})) = \text{rank}(\mathbf{U}_1^{(2)}) \leq \text{rank}(\mathbf{U}^{(2)}) = \dim(\text{Col}(\mathbf{U}^{(2)})) = d.$$

Since $\text{Col}(\mathbf{U}) = \text{Col}(\mathbf{U}^{(2)}) + \text{Col}(\mathbf{U}^{(1)})$, and $\text{Col}(\mathbf{U}^{(1)}) \subseteq \mathbf{L}_1$, the assumption $\text{Sp}(\mathbf{Y}) + \mathbf{L}_1 \supseteq \text{Col}(\mathbf{U})$ is equivalent to $\text{Sp}(\mathbf{Y}) + \mathbf{L}_1 \supseteq \text{Col}(\mathbf{U}^{(2)})$. Recall $\dim(\text{Sp}(\mathbf{Y})) \leq p$, and $\text{Col}(\mathbf{U}^{(2)})$ is a subspace in \mathbb{R}^{KD} that intersects \mathbf{L}_1 only at origin, we have $d \leq p$.

Apply $\text{Sp}(\mathbf{Y}) + \mathbf{L}_1 \supseteq \text{Col}(\mathbf{U}^{(2)})$ and $\dim(\text{Col}(\mathbf{U}^{(2)})) = d$, there exists $\mathbf{Y}_0 \in \mathbb{R}^{d \times D}$ and \mathbf{L} , a d -dimensional subspace in \mathbb{R}^p , such that

$$\text{Sp}(\mathbf{P}_L^\top [\mathbf{Y} - (\mathbf{Y}_0, \mathbf{Y}_0, \dots, \mathbf{Y}_0)]) = \text{Col}(\mathbf{U}^{(2)}).$$

Recall the property (27), there exists $\mathbf{Y}_i \in \mathbb{R}^{d \times d}$ for $1 \leq i \leq K$ and \mathbf{L}_0 , a d -dimensional subspace in \mathbb{R}^D that contains $\text{Sp}(\mathbf{U}_1^{(2)\top})$, such that

$$\mathbf{P}_L^\top [\mathbf{Y} - (\mathbf{Y}_0, \mathbf{Y}_0, \dots, \mathbf{Y}_0)] = (\mathbf{Y}_1 \mathbf{P}_{\mathbf{L}_0}^\top, \mathbf{Y}_2 \mathbf{P}_{\mathbf{L}_0}^\top, \dots, \mathbf{Y}_K \mathbf{P}_{\mathbf{L}_0}^\top). \quad (28)$$

In addition, since $\text{Col}(\mathbf{U}^{(2)})$ is orthogonal to \mathbf{L}_1 .

$$\mathbf{Y}_K = - \sum_{i=1}^{K-1} \mathbf{Y}_i. \quad (29)$$

Combining (28), (29), and the estimation $1 \leq d \leq p$, every \mathbf{Y} that does not satisfy (26) lies in the set $\cup_{d=1}^p \mathcal{M}_d$.

4.3.3. Step 3: counting the dimension of \mathcal{M}_d

In this step, we count the dimensions of \mathcal{M}_d for all $1 \leq d \leq p$, and show that they are smaller than pKD , the dimension of $\mathbb{R}^{p \times KD}$, which implies that generic $\mathbf{Y} \in \mathbb{R}^{p \times KD}$ does not belong to $\cup_{d=1}^p \mathcal{M}_d$, and (26) is proved.

For any d , the degree of freedom is $\{\mathbf{Y}_i\}_{i=1}^{K-1}$ is $(K-1)d^2$, the degree of freedom of \mathbf{Y}_0 is Dd , the degree of freedom of \mathbf{Z} is $KD(p-d)$, the degree of

freedom of \mathbf{L} and \mathbf{L}_1 are $d(p-d)$ and $d(D-d)$ respectively. Considering that $d \leq p \leq D-1$, the total dimension of \mathcal{M}_d is smaller than KDp , the dimension of $\mathbb{R}^{p \times KD}$. Since g_d is smooth and the dimension of its range is larger than its domain, all elements in \mathcal{M}_d are its critical values. Applying [17, Theorem 6.8], \mathcal{M}_d has measure zero in $\mathbb{R}^{(D-1) \times KD}$. Therefore, $\mathbb{R}^{(D-1) \times KD} \setminus \mathcal{M}_d$ is dense. Since $\mathbb{R}^{(D-1) \times KD} \setminus \mathcal{M}_d$ is a closed set, generic \mathbf{Y} does not lie in the set \mathcal{M}_d . Combining this result for all $1 \leq d \leq p$, generic \mathbf{Y} does not lie in the set $\cup_{d=1}^p \mathcal{M}_d$. \square

4.4. Proof of Lemma 4.1

We first state two lemmas that are rather easy to verify.

Lemma 4.2. *When $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times l}$, where $l \geq n$ and the singular values of \mathbf{B} are $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$, then*

$$\sigma_n \|\mathbf{A}\|_F \leq \|\mathbf{AB}\|_F \leq \sigma_1 \|\mathbf{A}\|_F. \quad (30)$$

Proof. First we claim that for any $\mathbf{x} \in \mathbb{R}^n$,

$$\sigma_n^2 \|\mathbf{x}\|^2 \leq \|\mathbf{x}^\top \mathbf{B}\|^2 \leq \sigma_1^2 \|\mathbf{x}\|^2. \quad (31)$$

Assuming that the SVD decomposition of \mathbf{B} is given by $\mathbf{B} = \mathbf{U}_B \text{diag}(\sigma_1, \dots, \sigma_n) \mathbf{V}_B^\top$, then the first inequality in (31) can be proved as follows:

$$\|\mathbf{x}^\top \mathbf{B}\|^2 = \mathbf{x}^\top \mathbf{B} \mathbf{B}^\top \mathbf{x} = \mathbf{x}^\top \mathbf{U}_B \text{diag}(\sigma_1^2, \dots, \sigma_n^2) \mathbf{U}_B^\top \mathbf{x} \geq \sigma_n^2 \|\mathbf{x}^\top \mathbf{U}_B\|^2 = \sigma_n^2 \|\mathbf{x}\|^2,$$

and second inequality in (31) can be proved similarly.

Assuming that $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_m)$, and combining (31) with $\mathbf{x} = \mathbf{a}_i$ for $1 \leq i \leq m$, (30) is proved. \square

Lemma 4.3. *The smallest eigenvalue of*

$$\begin{pmatrix} a & b \\ b & 0 \end{pmatrix}$$

is smaller than

$$-\frac{b^2}{\max(a, 0) + |b|}.$$

Proof. The smaller eigenvalue is

$$\frac{a - \sqrt{a^2 + 4b^2}}{2} = -\frac{2b^2}{a + \sqrt{a^2 + 4b^2}} \leq -\frac{2b^2}{a + (|a| + 2|b|)} = -\frac{b^2}{\max(a, 0) + |b|}.$$

\square

Proof of Lemma 4.1. First of all, WLOG we may assume that

$$\mathbf{Y} = \begin{pmatrix} \mathbf{I}_{M \times M} \\ \mathbf{0}_{M \times D-M} \end{pmatrix}. \quad (32)$$

If we proved the case (32), then other cases can be proved as follows. For generic \mathbf{Y} , there are invertible matrices $\mathbf{B} \in \mathbb{R}^{M \times M}$ and $\mathbf{A} \in \mathbb{R}^{D \times D}$ such that

$$\mathbf{B}\mathbf{Y}\mathbf{A} = \begin{pmatrix} \mathbf{I}_{M \times M} \\ \mathbf{0}_{M \times D-M} \end{pmatrix}.$$

Note that

$$(\mathbf{A}^\top \mathbf{L} \mathbf{B}^{-1})(\mathbf{B}\mathbf{Y}\mathbf{A}) + (\mathbf{B}\mathbf{Y}\mathbf{A})^\top (\mathbf{A}^\top \mathbf{L} \mathbf{B}^{-1})^\top = \mathbf{A}^\top (\mathbf{L}\mathbf{Y} + \mathbf{Y}^\top \mathbf{L}^\top) \mathbf{A},$$

we have

$$\lambda_{\min}((\mathbf{A}\mathbf{L}\mathbf{B}^{-1})(\mathbf{B}\mathbf{Y}\mathbf{A}) + (\mathbf{B}\mathbf{Y}\mathbf{A})^\top (\mathbf{A}\mathbf{L}\mathbf{B}^{-1})^\top) \geq -\|\mathbf{A}\|^2 \epsilon,$$

and similarly

$$\lambda_{\min}((\mathbf{A}\mathbf{L}\mathbf{B}^{-1})(\mathbf{B}\mathbf{Z}\mathbf{A}) + (\mathbf{B}\mathbf{Z}\mathbf{A})^\top (\mathbf{A}\mathbf{L}\mathbf{B}^{-1})^\top) \geq -\|\mathbf{A}\|^2 \epsilon.$$

Since the case (32) is assumed to be proved,

$$\|\mathbf{A}\mathbf{L}\mathbf{B}^{-1}\|_F \leq \epsilon \|\mathbf{A}\|^2 c(\mathbf{B}\mathbf{Y}\mathbf{A}, \mathbf{B}\mathbf{Z}\mathbf{A}).$$

Applying Lemma 4.2, $\|\mathbf{A}\mathbf{L}\mathbf{B}^{-1}\|_F \geq \|\mathbf{L}\|_F \sigma_{\min}(\mathbf{A}) / \|\mathbf{B}\|$, and the generic case is proved with $c(\mathbf{Y}, \mathbf{Z}) = \frac{\|\mathbf{A}\|^2 \|\mathbf{B}\|}{\sigma_{\min}(\mathbf{A})} c(\mathbf{B}\mathbf{Y}\mathbf{A}, \mathbf{B}\mathbf{Z}\mathbf{A})$, which is positive for generic \mathbf{Z} since $c(\mathbf{B}\mathbf{Y}\mathbf{A}, \mathbf{B}\mathbf{Z}\mathbf{A})$ is positive for generic $\mathbf{B}\mathbf{Z}\mathbf{A}$.

The rest of the proof will assume (32) and use induction on k , which is the integer such that $k(D-M) < M \leq (k+1)(D-M)$. For $k=1$, let us denote

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \end{pmatrix}, \quad \mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2),$$

where $\mathbf{L}_1, \mathbf{Z}_1 \in \mathbb{R}^{M \times M}$, $\mathbf{L}_2 \in \mathbb{R}^{M \times (D-M)}$, and $\mathbf{Z}_2 \in \mathbb{R}^{(D-M) \times M}$. Then

$$\mathbf{L}\mathbf{Y} + \mathbf{Y}^\top \mathbf{L}^\top = \begin{pmatrix} \mathbf{L}_1 + \mathbf{L}_1^\top & \mathbf{L}_2^\top \\ \mathbf{L}_2 & \mathbf{0} \end{pmatrix}$$

If $\|\mathbf{L}_2\| = b$ and $\|\mathbf{L}_1\| = a$, then there exists $\mathbf{u} \in \mathbb{R}^M, \mathbf{v} \in \mathbb{R}^{D-M}$ such that $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$ and $\mathbf{u}^\top \mathbf{L}_2 \mathbf{v} = b$. Then

$$\begin{pmatrix} \mathbf{u} & \mathbf{0} \\ \mathbf{0} & \mathbf{v} \end{pmatrix} (\mathbf{L}\mathbf{Y} + \mathbf{Y}^\top \mathbf{L}^\top) \begin{pmatrix} \mathbf{u} & \mathbf{0} \\ \mathbf{0} & \mathbf{v} \end{pmatrix}^\top = \begin{pmatrix} \mathbf{u}^\top (\mathbf{L}_1 + \mathbf{L}_1^\top) \mathbf{u} & b \\ b & \mathbf{0} \end{pmatrix},$$

and its smallest eigenvalue is larger than $-\epsilon$ since $\begin{pmatrix} \mathbf{u} & \mathbf{0} \\ \mathbf{0} & \mathbf{v} \end{pmatrix}^\top \begin{pmatrix} \mathbf{u} & \mathbf{0} \\ \mathbf{0} & \mathbf{v} \end{pmatrix} = \mathbf{I}$. Applying Lemma 4.3 with the estimation $\mathbf{u}^\top (\mathbf{L}_1 + \mathbf{L}_1^\top) \mathbf{u} \leq 2a$, we have

$$b^2 \leq \epsilon(2a + b). \tag{33}$$

In another aspect, applying Lemma 4.2 we have

For generic \mathbf{Z} , $C_1 \|\mathbf{L}_1\|_F \geq \|\mathbf{L}_1 \mathbf{Z}_2\|_F \geq C_2 \|\mathbf{L}_1\|_F$ for some $C_1, C_2 > 0$.

Therefore, we can find $\mathbf{u} \in \mathbb{R}^M$ and $\mathbf{v} \in \mathbb{R}^{N-M}$ such that $|\mathbf{u}^\top (\mathbf{L}_1 \mathbf{Z}_2) \mathbf{v}| \geq C_2 a$. Note that

$$\begin{pmatrix} \mathbf{u} & 0 \\ 0 & \mathbf{v} \end{pmatrix} (\mathbf{L} \mathbf{Z} + \mathbf{Z}^\top \mathbf{L}^\top) \begin{pmatrix} \mathbf{u} & 0 \\ 0 & \mathbf{v} \end{pmatrix}^\top = \begin{pmatrix} \mathbf{u}^\top (\mathbf{L}_1 \mathbf{Z}_1 + \mathbf{Z}_1^\top \mathbf{L}_1^\top) \mathbf{u} & \mathbf{u}^\top (\mathbf{Z}_2^\top \mathbf{L}_1^\top) \mathbf{v} \\ \mathbf{u}^\top (\mathbf{L}_1 \mathbf{Z}_2) \mathbf{v} & 0 \end{pmatrix} + \mathbf{L}',$$

where $\|\mathbf{L}'\| < C_3 b$. Since $|\mathbf{u}^\top (\mathbf{L}_1 \mathbf{Z}_1 + \mathbf{Z}_1^\top \mathbf{L}_1^\top) \mathbf{u}| \leq C_4 a$ for some $C_4 > 0$, Lemma 4.3 shows that the smallest eigenvalue of

$$\begin{pmatrix} \mathbf{u}^\top (\mathbf{L}_1 \mathbf{Z}_1 + \mathbf{Z}_1^\top \mathbf{L}_1^\top) \mathbf{u} & \mathbf{u}^\top (\mathbf{Z}_2^\top \mathbf{L}_1^\top) \mathbf{v} \\ \mathbf{u}^\top (\mathbf{L}_1 \mathbf{Z}_2) \mathbf{v} & 0 \end{pmatrix}$$

is smaller than

$$-\frac{C_2^2}{C_4 + C_1} a.$$

Let $C'_4 = \frac{C_2^2}{C_4 + C_1}$, applying $\lambda_{\min}(\mathbf{L} \mathbf{Z} + \mathbf{Z}^\top \mathbf{L}^\top) \geq -\epsilon$ we have

$$C_3 b - C'_4 a > -\epsilon. \quad (34)$$

This means $a < \frac{1}{C'_4} (C_3 b + \epsilon)$. Plug it into (33), we have

$$b^2 \leq \epsilon(b + 2a) \leq \epsilon \left(b + \frac{2}{C'_4} (C_3 b + \epsilon) \right) = \epsilon \left(1 + \frac{2C_3}{C'_4} \right) b + \frac{2}{C'_4} \epsilon^2,$$

which implies that $b < C\epsilon$. Applying (34), we also have $a < C'\epsilon$ and the Lemma is proved for $k = 1$.

For $k > 1$, let us first write

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_{1,1} & \mathbf{L}_{1,2} \\ \mathbf{L}_{2,1} & \mathbf{L}_{2,2} \end{pmatrix},$$

where $\mathbf{L}_{1,1} \in \mathbb{R}^{M \times (2M-D)}$, $\mathbf{L}_{1,2} \in \mathbb{R}^{M \times (D-M)}$, $\mathbf{L}_{2,1} \in \mathbb{R}^{(D-M) \times (2M-D)}$, $\mathbf{L}_{2,2} \in \mathbb{R}^{(D-M) \times (D-M)}$ and similarly write

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_{1,1} & \mathbf{Z}_{1,2} \\ \mathbf{Z}_{2,1} & \mathbf{Z}_{2,2} \end{pmatrix},$$

where $\mathbf{Z}_{1,1} \in \mathbb{R}^{(2M-D) \times M}$, $\mathbf{Z}_{1,2} \in \mathbb{R}^{(D-M) \times M}$, $\mathbf{Z}_{2,1} \in \mathbb{R}^{(2M-D) \times (D-M)}$, $\mathbf{Z}_{2,2} \in \mathbb{R}^{(D-M) \times (D-M)}$. WLOG we may assume that $\mathbf{Z}_{1,2} = \mathbf{0}$, by finding an appropriate orthogonal matrix $\mathbf{A} \in \mathbb{R}^{M \times M}$ and consider $(\mathbf{L} \mathbf{A}', \mathbf{A} \mathbf{Y}, \mathbf{A} \mathbf{Y})$ instead of $(\mathbf{L}, \mathbf{Y}, \mathbf{Z})$, then

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_{1,1} & \mathbf{0} \\ \mathbf{Z}_{2,1} & \mathbf{Z}_{2,2} \end{pmatrix},$$

and for generic \mathbf{Z} , $\mathbf{Z}_{2,2}$ is full-rank, i.e., the rank is $D-M$. Let $b = \|(\mathbf{L}_{2,1}, \mathbf{L}_{2,2})\|_F$, $a_1 = \|\mathbf{L}_{1,1}\|_F$ and $a_2 = \|\mathbf{L}_{1,2}\|_F$. Then following the proof of (33), applying $\lambda_{\min}(\mathbf{L}\mathbf{Y} + \mathbf{Y}^\top \mathbf{L}^\top) \geq -\epsilon$ we have

$$\frac{b^2}{a_1 + a_2 + b} \leq \epsilon. \quad (35)$$

If another aspect,

$$\mathbf{L}\mathbf{Z} + \mathbf{Z}^\top \mathbf{L}^\top = \begin{pmatrix} \mathbf{L}_{1,1}\mathbf{Z}_{1,1} + \mathbf{L}_{1,2}\mathbf{Z}_{2,1} + \mathbf{Z}_{1,1}^\top \mathbf{L}_{1,1}^\top + \mathbf{Z}_{1,2}^\top \mathbf{L}_{2,1}^\top & \mathbf{L}_{1,2}\mathbf{Z}_{2,2} \\ \mathbf{Z}_{2,2}^\top \mathbf{L}_{1,2}^\top & \mathbf{0} \end{pmatrix} + \mathbf{L}',$$

where $\|\mathbf{L}'\| < C_1 b$. Again Lemma 4.2 means that we have $C_2 a_2 \leq \|\mathbf{L}_{1,2}\mathbf{Z}_{2,2}\| \leq C_3 a_2$ and $\|\mathbf{L}_{1,1}\mathbf{Z}_{1,1} + \mathbf{L}_{1,2}\mathbf{Z}_{2,1} + \mathbf{Z}_{1,1}^\top \mathbf{L}_{1,1}^\top + \mathbf{Z}_{1,2}^\top \mathbf{L}_{2,1}^\top\| \leq C_4(a_1 + a_2)$. Lemma 4.3 then implies

$$\frac{C_2^2 a_2^2}{C_4(a_1 + a_2) + C_3 a_2} \leq \epsilon + C_1 b. \quad (36)$$

At last, note that $\mathbf{L}_{1,1}\mathbf{Z}_{1,1} + \mathbf{L}_{1,2}\mathbf{Z}_{2,1} + \mathbf{Z}_{1,1}^\top \mathbf{L}_{1,1}^\top + \mathbf{Z}_{1,2}^\top \mathbf{L}_{2,1}^\top$ is a submatrix of $\mathbf{L}\mathbf{Z} + \mathbf{Z}^\top \mathbf{L}^\top$, so its smallest eigenvalue is also larger than $-\epsilon$. Since there exists C_5 such that $\|\mathbf{L}_{1,2}\mathbf{Z}_{2,1} + \mathbf{Z}_{1,2}^\top \mathbf{L}_{2,1}^\top\| \leq C_5 a_2$, the smallest eigenvalue of $\mathbf{L}_{1,1}\mathbf{Z}_{1,1} + \mathbf{Z}_{1,1}^\top \mathbf{L}_{1,1}^\top$ is larger than $-\epsilon - C_5 a_2$. Let

$$\mathbf{Y}_{1,1} = \begin{pmatrix} \mathbf{I}_{(2M-D) \times (2M-D)} \\ \mathbf{0}_{(2M-D) \times (D-M)} \end{pmatrix}$$

and using the same argument, the smallest eigenvalue of $\mathbf{L}_{1,1}\mathbf{Z}_{1,1} + \mathbf{Z}_{1,1}^\top \mathbf{L}_{1,1}^\top$ is larger than $-\epsilon - C_5' a_2$. Since $(k-1)(D-M) < 2M-D \leq k(D-M)$, we may apply the case $k-1$ to $\mathbf{L}_{1,1}$, $\mathbf{Y}_{1,1}$ and $\mathbf{Z}_{1,1}$ and have

$$a_1 \leq C_6((C_5 + C_5')a_2 + \epsilon). \quad (37)$$

Plug in (37) to (35) and (36) we have

$$b^2 \leq \epsilon C_7(\epsilon + a_2 + b) \quad (38)$$

and

$$a_2^2 \leq C_8(\epsilon + a_2)(\epsilon + b), \text{ i.e., } a_2(a_2 - C_8 b) \leq \epsilon C_8(\epsilon + a_2 + b). \quad (39)$$

If $a_2 \geq 2C_8 b$, then (39) implies

$$\frac{1}{2}a_2^2 \leq \epsilon C_8(\epsilon + a_2 + \frac{1}{2C_8}a_2),$$

which implies $a_2 \leq C\epsilon$. This then implies $b \leq C'\epsilon$ (from assumption) and $a_1 \leq C''\epsilon$ (from (37)).

If $a_2 < 2C_8 b$, then (38) implies $b < C\epsilon$, which then implies $a_1, a_2 < C'\epsilon$.

For either case we have $a_1 + a_2 + b < C''' \epsilon$, and Lemma 4.1 is proved since $\|\mathbf{L}\|_F \leq \|\mathbf{L}_{1,1}\|_F + \|\mathbf{L}_{1,2}\|_F + \|(\mathbf{L}_{2,1}, \mathbf{L}_{2,2})\|_F = a + b + c$. \square

Acknowledgment

The authors were partially supported by Award Number R01GM090200 from the NIGMS, FA9550-12-1-0317 and FA9550-13-1-0076 from AFOSR, LTR DTD 06-05-2012 from the Simons Foundation, and the Moore Foundation Data-Driven Discovery Investigator Award. The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

References

References

- [1] J. Gower, G. Dijksterhuis, Procrustes Problems, Oxford Statistical Science Series.
- [2] A. S. Bandeira, C. Kennedy, A. Singer, Approximating the little Grothendieck problem over the orthogonal and unitary groups, *Mathematical Programming* (2016) 1–43.
- [3] A. Nemirovski, Sums of random symmetric matrices and quadratic optimization under orthogonality constraints, *Mathematical Programming* 109 (2-3) (2007) 283–317.
- [4] M. X. Goemans, D. P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, *J. ACM* 42 (6) (1995) 1115–1145.
- [5] J. Gower, Generalized procrustes analysis, *Psychometrika* 40 (1) (1975) 33–51.
- [6] J. Ten Berge, Orthogonal procrustes rotation for two or more matrices, *Psychometrika* 42 (2) (1977) 267–276.
- [7] A. Shapiro, J. D. Botha, Dual algorithm for orthogonal procrustes rotations, *SIAM Journal on Matrix Analysis and Applications* 9 (3) (1988) 378–383.
- [8] A. S. Bandeira, C. Kennedy, A. Singer, Approximating the little grothendieck problem over the orthogonal and unitary groups, *Mathematical Programming* 160 (1) (2016) 433–475.
- [9] Z. Kam, The reconstruction of structure from electron micrographs of randomly oriented particles, *Journal of Theoretical Biology* 82 (1) (1980) 15 – 39.
- [10] T. Bhamre, T. Zhang, A. Singer, Orthogonal matrix retrieval in cryo-electron microscopy, 12th IEEE International Symposium on Biomedical Imaging (ISBI 2015).

- [11] A. M.-C. So, J. Zhang, Y. Ye, On approximating complex quadratic optimization problems via semidefinite programming relaxations, *Math. Program.* 110 (1) (2007) 93–110.
- [12] A.-C. So, Moment inequalities for sums of random matrices and their applications in optimization, *Mathematical Programming* 130 (1) (2011) 125–151.
- [13] A. Naor, O. Regev, T. Vidick, Efficient rounding for the noncommutative grothendieck inequality, in: *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing, STOC '13*, 2013, pp. 71–80.
- [14] J. B. Keller, Closest unitary, orthogonal and Hermitian operators to a given operator, *Mathematics Magazine* 48 (4) (1975) 192–197.
- [15] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1, <http://cvxr.com/cvx> (Mar. 2014).
- [16] H. R. Tütüncü, C. K. Toh, J. M. Todd, Solving semidefinite-quadratic-linear programs using sdpt3, *Mathematical Programming* 95 (2) (2003) 189–217.
- [17] J. Lee, *Introduction to Smooth Manifolds*, Graduate Texts in Mathematics, Springer, 2003.