Explorations on Procrustes Problem with $2, \infty$ Norm

Yuqi Tu

April 2024

Abstract

This report focuses on the Procrustes problem with the $2, \infty$ norm, i.e., finding the optimal orthonormal matrix $\mathbf{W}^*_{2,\infty}$ which achieves the minimal value of $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$ where $\mathbf{X}$, $\mathbf{Y}$ are given orthogonal matrices. We investigate the computation of $\mathbf{W}^*_{2,\infty}$ and explore key properties of the norm $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$. We begin by addressing a simplified case, which leads to an algorithm for determining $\mathbf{W}^*_{2,\infty}$, followed by empirical results comparing $\mathbf{W}^*_{2,\infty}$ and $\mathbf{W}^*_F$ (the optimal matrix solving Procrustes problem with Frobenius norm) across different dimensions of $\mathbf{X}$ and $\mathbf{Y}$. These results motivate a theoretical investigation into various bounds for $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$, particularly when $\mathbf{Y}$ is a delocalized matrix.

For $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{4\times 2}$, where $\mathbf{Y}$ is delocalized, we provide a theoretical lower bound for the unknown upper bound of $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$ in terms of $\|\sin\Theta(\mathbf{X}, \mathbf{Y})\|_2$. As the dimensionality increases, the behavior of this norm becomes more complex due to concentration in high dimensions. While an exact upper bound remains elusive, we explore several key properties, including the maximal value of the norm.

To analyze properties in high dimensions further, instead of generating $\mathbf{X}$ from the Stiefel manifold, we assume that the entries of $\mathbf{X}$ are drawn from an i.i.d. Gaussian distribution. We leverage the properties of high-dimensional Gaussian vectors to derive several results, including a high-probability bound for $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$ and an upper bound for the expectation of this norm, whether or not $\mathbf{Y}$ is delocalized.

1 Introduction

1.1 Problem Set Up

Suppose we have two matrices, $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m\times n}$ and we want to solve

$$\operatorname*{arginf}_{\mathbf{W}\in\mathbb{O}^{n\times n}} \|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_\eta,$$

where η denotes the norm we choose. This problem is called Procrustes Problem which requires us to solve for $\mathbf{W}$ as well as $\|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_\eta$.

Notations: For any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, the two-to-infinity norm is given by $\|\mathbf{A}\|_{2,\infty} = \sup_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_\infty$ where $\mathbf{x} \in \mathbb{R}^n$, and the Frobenius norm is given by $\|\mathbf{A}\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$. For simplicity, we define $\mathbf{W}^*_F \coloneqq \operatorname{arginf}_{\mathbf{W}\in\mathbb{O}^{n\times n}} \|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_F$ and $\mathbf{W}^*_{2,\infty} \coloneqq \operatorname{arginf}_{\mathbf{W}\in\mathbb{O}^{n\times n}} \|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_{2,\infty}$.

1.2 Overview and Motivations

Previous literature has found the solution to the Procrustes problem when η is the Frobenius norm. The minimizer $\mathbf{W}^*_F$ can be found by the singular value decomposition.

Theorem 1.1 (Procrustes Problem for Frobenius Norm) *Denote the singular value decomposition of $\mathbf{Y}^T\mathbf{X}$ as $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Then, $\mathbf{W}^*_F = \mathbf{U}\mathbf{V}^T$.*

Proof: See appendix. □

However, when η is the $2, \infty$ norm, finding the exact value of $\mathbf{W}^*_{2,\infty}$ is complicated and even intractable. Previous work [Cap20] use $\mathbf{W}^*_F$ to be a surrogate for $\mathbf{W}^*_{2,\infty}$ when $\mathbf{X}$ and $\mathbf{Y}$ have orthonormal columns. Several bounds on $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_F\|_{2,\infty}$ are proposed.

Definition 1.1 (Canonical Angles) *Given two matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m\times n}$ with orthonormal columns representing two n-dimensional subspaces of $\mathbb{R}^m$, the canonical angles between the subspaces are defined as follows. Let $\sigma_i(\mathbf{Y}^\top\mathbf{X})$ be the singular values of $\mathbf{Y}^\top\mathbf{X}$, ordered in non-increasing order. The canonical angles are given by:*

$$\Theta(\mathbf{X},\mathbf{Y}) := \operatorname{diag}\left(\cos^{-1}\left(\sigma_1\left(\mathbf{Y}^\top\mathbf{X}\right)\right), \cos^{-1}\left(\sigma_2\left(\mathbf{Y}^\top\mathbf{X}\right)\right), \ldots, \cos^{-1}\left(\sigma_n\left(\mathbf{Y}^\top\mathbf{X}\right)\right)\right).$$

Theorem 1.2 (Corollary 4.3 in [Cap20]) *Let $\mathbf{X}, \mathbf{Y} \in \mathbb{O}_{m,n}$, $\alpha(a) := \sqrt{2-2\sqrt{1-a^2}}$, $\beta(b) := 1-\sqrt{1-b^2}$, then*

$$\frac{1}{\sqrt{m}}\alpha\left(\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2\right) \le \|\mathbf{X}-\mathbf{Y}\mathbf{W}_F^*\|_{2,\infty} \le \alpha\left(\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2\right).$$

Theorem 1.3 (Theorem 4.4 in [Cap20]) *Let $\mathbf{X}, \mathbf{Y} \in \mathbb{O}_{m,n}$, and $\mathbf{W}_F^* \in \mathbb{O}_n$. Then for $\mathbf{Y}_\perp \in \mathbb{O}_{m,m-n}$ such that $[\mathbf{Y} \mid \mathbf{Y}_\perp] \in \mathbb{O}_m$,*

$$\|\mathbf{X}-\mathbf{V}\mathbf{W}_F^*\|_{2,\infty} \le \|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2\,\|\mathbf{Y}_\perp\|_{2,\infty} + \beta\left(\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2\right)\|\mathbf{Y}\|_{2,\infty}$$

Definition 1.2 (Delocalized Matrix) *We call a matrix $\in \mathbb{R}^{m\times 2}$ is a canonical two-block balanced delocalized matrix (in this report, for simplicity, we call it delocalized) if such matrix is of the format of*

$$\begin{pmatrix} \frac{1}{\sqrt{m}} & \frac{1}{\sqrt{m}} \\ \vdots & \vdots \\ \frac{1}{\sqrt{m}} & \frac{1}{\sqrt{m}} \\ \frac{1}{\sqrt{m}} & -\frac{1}{\sqrt{m}} \\ \vdots & \vdots \\ \frac{1}{\sqrt{m}} & -\frac{1}{\sqrt{m}} \end{pmatrix}.$$

Lemma 1.4 (Delocalized version of Theorem 1.3) *If we have $\mathbf{V}$ to be a delocalzied matrix following the definition above, then*

$$\|\mathbf{X}-\mathbf{Y}\mathbf{W}_F^*\|_{2,\infty} \le \|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2\sqrt{\frac{m-2}{m}} + \beta\left(\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2\right)\sqrt{\frac{2}{m}} \tag{1}$$

Proof: This follows by Theorem 1.3 since here $\|\mathbf{Y}_\perp\|_{2,\infty} = \sqrt{\frac{m-2}{m}}$, $\|\mathbf{Y}\|_{2,\infty} = \sqrt{\frac{2}{m}}$. $\square$

Motivated by the findings in [Cap20], this report focuses on determining the exact value of $\mathbf{W}_{2,\infty}^*$ and exploring several properties of the norm $\|\mathbf{X}-\mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$ when $\mathbf{X}, \mathbf{Y} \in \mathbb{O}^{m\times n}$. In section 2.1, we begin by analyzing a simplified case, which leads to the development of an algorithm for computing $\mathbf{W}_{2,\infty}^*$. Using this algorithm, we present some empirical results in section 2.2.

Building on these empirical findings, in section 3, we investigate potential upper bounds in terms of $\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2$ for the case where $\mathbf{X}, \mathbf{Y} \in \mathbb{O}^{4\times 2}$, with particular attention to the scenario in which $\mathbf{Y}$ is a delocalized matrix. However, as dimensionality increases, the behavior of the norm $\|\mathbf{X}-\mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$ becomes more complex due to the concentration of measure in higher dimensions, which necessitates additional simulations to see the whole distributions. While deriving a precise upper bound remains challenging, we are able to identify key properties, such as the lower bound for maximal value of $\|\mathbf{X}-\mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$ when $\mathbf{Y}$ is delocalized (section 4).

For higher dimensions, instead of generating $\mathbf{X}$ and $\mathbf{Y}$ from the Stiefel manifold, we sample their entries from $\mathcal{N}(0, \frac{1}{m})$. Utilizing the near-orthogonality of Gaussian vectors and their near-unit norms in high dimensions, we derive similar results. Additionally, the properties of Gaussian distributions enable us to identify certain symmetries, leading to several bounds: (1) When $\mathbf{Y}$ is delocalized, we establish an upper bound for $\mathbb{E}[\|\mathbf{X}-\mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}]$ (section 5.1);(2) We propose a high-probability bound leveraging the symmetry of high-dimensional spaces (section 5.3); and (3) We generalize the upper bound for the expectation of $\|\mathbf{X}-\mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$ when $\mathbf{Y}$ is no longer delocalized (section 5.4).

In the final section, we discuss several directions previously explored that did not yield promising results, offering insights into possible future research approaches.

2 For $\mathbf{W}$ to be a 2×2 matrix

In this section, we focus on the case where the matrix $\mathbf{W}$ is 2×2. We start with the simplest scenario, demonstrating how to mathematically compute $\mathbf{W}^*_{2,\infty}$ for given matrices $\mathbf{X}$ and $\mathbf{Y} \in \mathbb{R}^{2\times 2}$. Building on this, we extend the approach to develop an algorithm for computing $\mathbf{W}^*_{2,\infty}$ when $\mathbf{X}$ and $\mathbf{Y} \in \mathbb{R}^{m\times 2}$ for general $m \geq 2$. Using this algorithm, we present several visual comparisons of the Procrustes problem under different norms.

2.1 Simple Case: $\mathbf{W} \in \mathbb{O}^{2\times 2}$

We begin with a simple case where $\mathbf{X}, \mathbf{Y}$ and $\mathbf{W}$ are all 2×2 matrices. When $\mathbf{W}$ is a 2×2 orthogonal matrix, it can only be written as a rotation matrix or a reflection matrix, which allows us to use one single parameter θ and an additional parameter $\xi \in \{-1, 1\}$ (decides $\mathbf{W}$ is a rotation matrix or a reflection matrix) to represent $\mathbf{W}$. We write:

$$\mathbf{X} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} e & f \\ g & h \end{pmatrix},$$

$$\mathbf{W}_{\text{rot}} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad \mathbf{W}_{\text{ref}} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

Here, $\mathbf{W}_{\text{rot}}$ represents the rotation matrix, and $\mathbf{W}_{\text{ref}}$ represents the reflection matrix.

In the following, we focus on rotation matrix, and the case of reflection matrix can also be solved in a similar way.

We can convert previous problem into

$$\operatorname*{arginf}_{\theta\in[-\pi,\pi]} \|\mathbf{X} - \mathbf{Y}\mathbf{W}_{rot}\|_{2,\infty}.$$

We have

$$\mathbf{X} - \mathbf{Y}\mathbf{W}_{rot} = \begin{pmatrix} a - (e\cos\theta + f\sin\theta) & b - (-e\sin\theta + f\cos\theta) \\ c - (g\cos\theta + h\sin\theta) & d - (-g\sin\theta + h\cos\theta) \end{pmatrix}.$$

We denote the squared l_2 norm of each row as

$$\Phi(\theta) = (a - (e\cos\theta + f\sin\theta))^2 + (b - (-e\sin\theta + f\cos\theta))^2,$$

$$\Psi(\theta) = (c - (g\cos\theta + h\sin\theta))^2 + (d - (-g\sin\theta + h\cos\theta))^2.$$

Thus we have

$$\Phi(\theta) = a^2 + b^2 + e^2 + f^2 + 2(be - af)\sin\theta - 2(ae + bf)\cos\theta, \tag{2}$$
$$\Psi(\theta) = c^2 + d^2 + g^2 + h^2 + 2(dg - ch)\sin\theta - 2(cg + dh)\cos\theta. \tag{3}$$

We want to solve

$$\min_\theta \max\left(\Phi(\theta), \Psi(\theta)\right).$$

Let $H(\theta) = \Phi(\theta) - \Psi(\theta)$, and notice $\sin\theta = \frac{2x}{1+x^2}, \cos\theta = \frac{1-x^2}{1+x^2}$ where $x = \tan\frac{\theta}{2}$. Now we consider solving

$$H(x) = 0,$$

since this gives the boundary values for $H(\theta) \geq 0$ and $H(\theta) < 0$. For simplicity, we denote

$$\Phi(\theta) = A + 2B\sin\theta - 2C\cos\theta, \tag{4}$$
$$\Psi(\theta) = D + 2E\sin\theta - 2F\cos\theta. \tag{5}$$

Then $H(x)$ becomes

$$(A - D + 2C - 2F)x^2 + 4(B - E)x + (A - D - 2C + 2F) = 0,$$

and

$$\Delta = (4(B - E))^2 - 4((A - D)^2 - (2C - 2F)^2).$$

(i) If $\Delta < 0$, then $\Phi(\theta) \geq \Psi(\theta)$ or $\Psi(\theta) \geq \Phi(\theta)$ for any θ. This is the same as for any θ, the ℓ_2 norm of one row is always equal or greater than the other row. We denote the $\theta_1 := \operatorname{argmin}_\theta \Phi(\theta)$ and $\theta_2 := \operatorname{argmin}_\theta \Psi(\theta)$. In this case, the final answer we are looking for is given by $\min\{\Phi(\theta_1), \Psi(\theta_2)\}$.

(ii) If $\Delta \geq 0$, then $\Psi(\theta) = \Phi(\theta)$ for some θ. In this case, we need to calculate the θ when $\Psi(\theta) = \Phi(\theta)$. We denote $S = \{\theta | \Psi(\theta) = \Phi(\theta)\}$. And the final answer is given by $\min\{\Phi(\theta_1), \Psi(\theta_2)\, \Psi_{\theta \in S}(\theta)\}$.

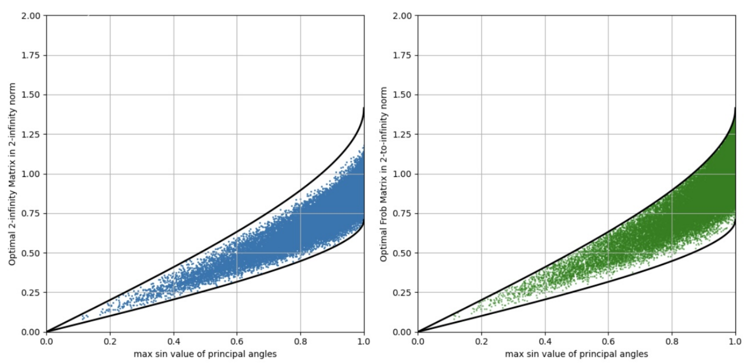
We can apply a similar strategy to find optimal reflection $\mathbf{W}_{ref}$, and the final answer is chosen from the minimal of rotation and reflection case.

When $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m \times 2}$, the procedure remains the same, with the only difference being the need to compare more pairs of rows. This extension enables the development of an algorithm, which can be implemented in Python. The corresponding code is provided in the appendix.

2.2 Comparisons between $\|\mathbf{X} - \mathbf{YW}\|_{2,\infty}$ and $\|\mathbf{X} - \mathbf{YW}\|_F$

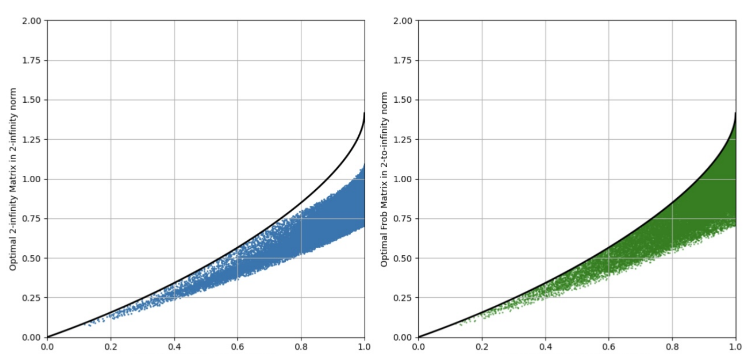
With the algorithm to find $\mathbf{W}^*_{2,\infty}$, we present some comparisons for different norms. In this section, we generate $\mathbf{X}, \mathbf{Y}$ following a similar way in [Cap20], i.e., we generate the entries of $\mathbf{X}$ and $\mathbf{Y}$ from $\mathcal{N}(0, 1)$ and then using Gram$-$Schmidt orthonormalization.

2.2.1 When $\mathbf{X}, \mathbf{Y} \in \mathbb{O}^{4 \times 2}$



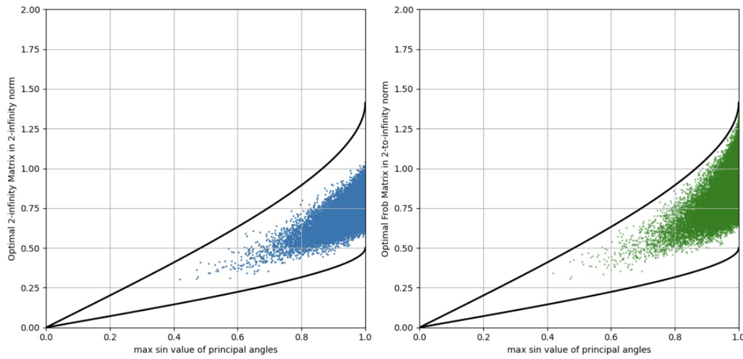
We generate 40000 pairs of $\mathbf{X}$ and $\mathbf{Y}$, and plot $\|\mathbf{X} - \mathbf{YW}^*_{2,\infty}\|_{2,\infty}$ using blue points and $\|\mathbf{X} - \mathbf{YW}^*_F\|_{2,\infty}$ using green points. The black curves are the bounds in Theorem 1.2 and Theorem 1.3.

2.2.2 When $\mathbf{X}, \mathbf{Y} \in \mathbb{O}^{4 \times 2}$ and $\mathbf{Y}$ is a delocalized matrix



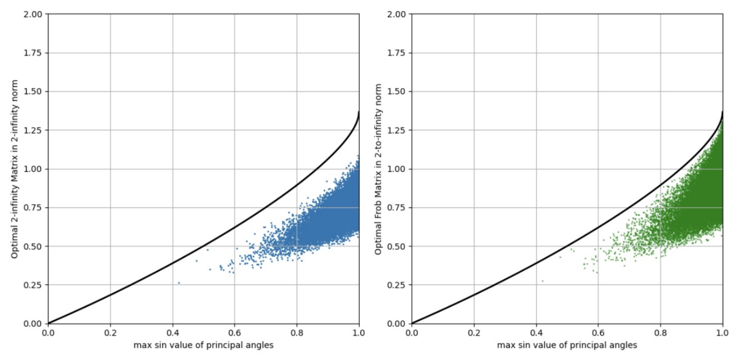
We generate 40000 $\mathbf{X}$ and $\mathbf{Y}$ is a fixed delocalized matrix, and plot $\|\mathbf{X} - \mathbf{YW}^*_{2,\infty}\|_{2,\infty}$ using blue points and $\|\mathbf{X} - \mathbf{YW}^*_F\|_{2,\infty}$ using green points. The black curve is the bound in Lemma 1.4.

2.2.3 When $\mathbf{X}, \mathbf{Y} \in \mathbb{O}^{8\times 2}$



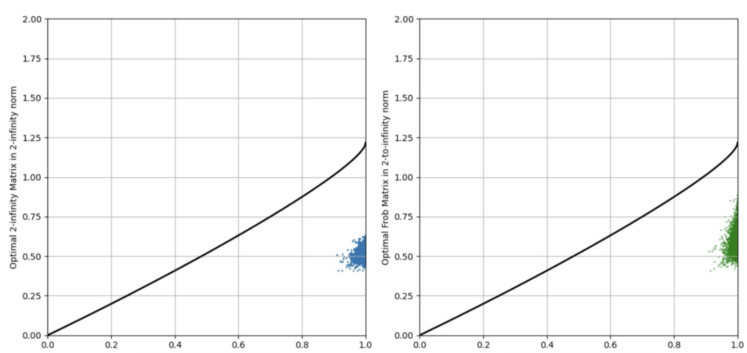
We generate 40000 pairs of $\mathbf{X}$ and $\mathbf{Y}$, and plot $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$ using blue points and $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{F}\|_{2,\infty}$ using green points. The black curves are the bounds in Theorem 1.2 and Theorem 1.3.

2.2.4 $\mathbf{X}, \mathbf{Y} \in \mathbb{O}^{8\times 2}$ and $\mathbf{Y}$ is a delocalized matrix



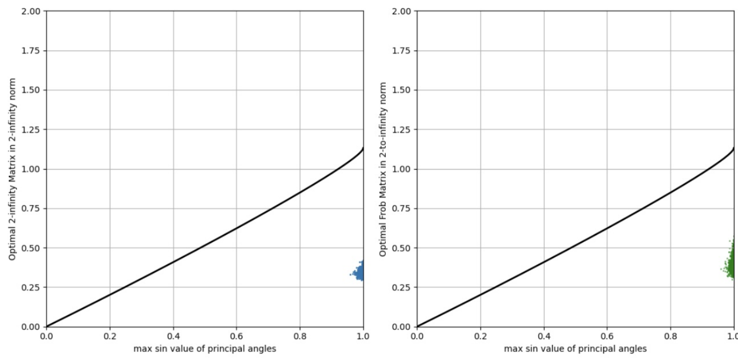
We generate 40000 $\mathbf{X}$ and $\mathbf{Y}$ is a fixed delocalized matrix, and plot $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$ using blue points and $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{F}\|_{2,\infty}$ using green points. The black curve is the bound in Lemma 1.4.

2.2.5 Higher Dimension, When $\mathbf{X}, \mathbf{Y} \in \mathbb{O}^{32\times 2}$ and $\mathbf{Y}$ is a delocalized matrix



We generate 4000 $\mathbf{X}$ and $\mathbf{Y}$ is a fixed delocalized matrix, and plot $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$ using blue points and $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{F}\|_{2,\infty}$ using green points. The black curve is the bound in Lemma 1.4.

2.2.6 Higher Dimension, When $\mathbf{X}, \mathbf{Y} \in \mathbb{O}^{100\times 2}$ and $\mathbf{Y}$ is a delocalized matrix



We generate 4000 $\mathbf{X}$ and $\mathbf{Y}$ is a fixed delocalized matrix, and plot $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$ using blue points and $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_F\|_{2,\infty}$ using green points. The black curve is the bound in Lemma 1.4.

Remark 1 *As the dimensionality increases, the points become more concentrated and tend to shift to the right. To capture the full picture in higher dimensions, a greater number of simulation iterations is required.*

3 Upper Bound for $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$, $\mathbf{X} \in \mathbb{O}^{4\times 2}$ and $\mathbf{Y}$ is delocalized

We have observed that certain upper bounds for $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_F\|_{2,\infty}$ also provide immediate upper bounds for $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$. But can we have a better approximation towards the unknown upper bound for $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$? In this section, we aim to refine our understanding of the unknown upper bound for $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$. We will focus on the case in which $\mathbf{X}$ and $\mathbf{Y} \in \mathbb{O}^{4\times 2}$ and $\mathbf{Y}$ is delocalized. The matrix products of $\mathbf{Y}$ and $\mathbf{W}$ are the following:

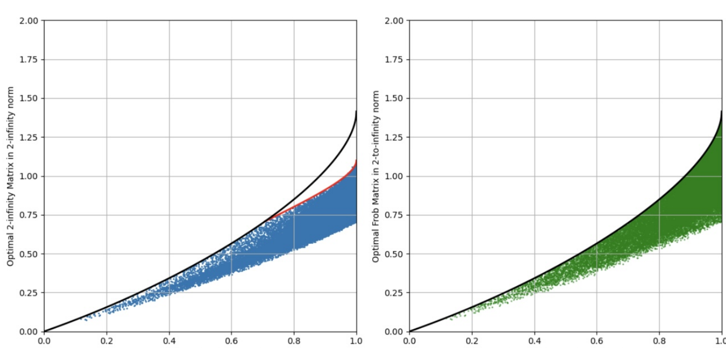
$$\mathbf{Y}\mathbf{W}_{rot} = \begin{pmatrix} \frac{\cos\theta+\sin\theta}{2} & \frac{\cos\theta-\sin\theta}{2} \\ \frac{\cos\theta+\sin\theta}{2} & \frac{\cos\theta-\sin\theta}{2} \\ \frac{\cos\theta-\sin\theta}{2} & \frac{-\cos\theta-\sin\theta}{2} \\ \frac{\cos\theta-\sin\theta}{2} & \frac{-\cos\theta-\sin\theta}{2} \end{pmatrix} \quad \text{and} \quad \mathbf{Y}\mathbf{W}_{ref} = \begin{pmatrix} \frac{\cos\theta+\sin\theta}{2} & \frac{-\cos\theta+\sin\theta}{2} \\ \frac{\cos\theta+\sin\theta}{2} & \frac{-\cos\theta+\sin\theta}{2} \\ \frac{\cos\theta-\sin\theta}{2} & \frac{\cos\theta+\sin\theta}{2} \\ \frac{\cos\theta-\sin\theta}{2} & \frac{\cos\theta+\sin\theta}{2} \end{pmatrix}.$$

By Lemma 1.4, we have

$$\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_F\|_{2,\infty} \leq \frac{1}{\sqrt{2}} \left(\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2 + 1 - \sqrt{1 - \|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2^2} \right). \tag{6}$$

We plot this bound in the figures below, and the bound is the black curve in the figures on the left hand side. The blue points on the left are pairs of $(\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2, \|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty})$ and the green points on the right are the pairs of $(\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2, \|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{\mathbf{F}}\|_{2,\infty})$.

When $\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2 \leq \frac{\sqrt{2}}{2}$, the upper bound for $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{\mathbf{F}}\|_{2,\infty}$ (6) meets the empirical upper bound (see the upper boundary of blue points in the figure on the left) for $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$. By the definition $\mathbf{W}^*_{2,\infty}$, necessarily $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty} \leq \|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_F\|_{2,\infty}$.



When $\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2 > \frac{\sqrt{2}}{2}$, we see there is a gap between $\|\mathbf{X}-\mathbf{Y}\mathbf{W}_F^*\|_{2,\infty}$ (6) and empirical upper bound for $\|\mathbf{X}-\mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$ (see the black curve and the upper boundary of the blue points in the top-left figure).

For this part, we propose a lower bound for the upper bound of $\|\mathbf{X}-\mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$ (see the red curve in the top-left figure). We demonstrate that there exist matrices that lie along this red curve, thus the upper bound for $\|\mathbf{X}-\mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$ cannot be smaller than this. Empirically, the lower bound and the true upper bound appear to be very close.

Claim 3.1 *We state there is a lower bound for the upper bound for $\|\mathbf{X}-\mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$ as the following:*

$$B(\mathbf{X},\mathbf{Y}) = \sqrt{\frac{\sqrt{2}}{2}\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2 + \frac{1}{2} - \frac{\sqrt{2}}{2}\sqrt{1-\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2^2}}, \qquad \textit{if} \quad \|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2 > \frac{\sqrt{2}}{2}.$$

Proof: To prove this, we show that we can recover all the matrices satisfying $B(\mathbf{X},\mathbf{Y})$ which have the corresponding $\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2$ and $\|\mathbf{X}-\mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$ value.

We give the proof by constructing special matrix, and we define these special matrices to be

$$\mathbf{X}_{sp} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ \cos\alpha & 0 \\ \sin\alpha & 0 \end{pmatrix},$$

where $\alpha \in [0, 2\pi]$. In such setting, we give the exact value of $B(\mathbf{X}_{sp},\mathbf{Y})$ and $\|\mathbf{X}_{sp}-\mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$. First, by calculation, the SVD for $\mathbf{Y}^T\mathbf{T}\mathbf{X}_{sp}$ will give $\mathbf{Y}^T\mathbf{T}\mathbf{X}_{sp} = \mathbf{U}_{sp}\mathbf{\Sigma}_{sp}\mathbf{V}_{sp}^T$ where

$$\mathbf{\Sigma}_{sp} = \begin{pmatrix} \frac{|\cos\alpha+\sin\alpha|}{\sqrt{2}} & 0 \\ 0 & \frac{\sqrt{2}}{2} \end{pmatrix}.$$

So the corresponding $\|\sin\Theta(\mathbf{X}_{sp},\mathbf{Y})\|_2$ have the following properties:

$$\|\sin\Theta(\mathbf{X}_{sp},\mathbf{Y})\|_2 = \begin{cases} \frac{\sqrt{2}}{2} & \text{if } |\cos\alpha+\sin\alpha| \geq 1, \\ \frac{|\cos\alpha-\sin\alpha|}{\sqrt{2}} & \text{otherwise.} \end{cases} \tag{7}$$

For case 1 in eq.(7), this corresponds to the case with $\|\sin\Theta(\mathbf{X}_{sp},\mathbf{Y})\|_2 \leq \frac{\sqrt{2}}{2}$ and we already have a good approximation for upper bound in this case. In the following, we show that case 2 in eq.(7) contains the matrices we will use to recover $B(\mathbf{X},\mathbf{Y})$. Without loss of generality, we only discuss the case in which $|\sin\alpha| \geq \frac{\sqrt{2}}{2} \geq |\cos\alpha|$, i.e., $\alpha \in [\frac{\pi}{2}, \frac{3\pi}{4}] \bigcup [\frac{3\pi}{2}, \frac{7\pi}{4}]$.

1. $\alpha \in [\frac{\pi}{2}, \frac{3\pi}{4}]$

   By the properties of trigonometric functions, $|\sin\alpha - \frac{\sqrt{2}}{2}| \leq |\cos\alpha + \frac{\sqrt{2}}{2}| \leq 1 - \frac{\sqrt{2}}{2} \leq \frac{\sqrt{2}}{2} \leq |\cos\alpha - \frac{\sqrt{2}}{2}| \leq 1 \leq |\sin\alpha + \frac{\sqrt{2}}{2}| \leq 1 + \frac{\sqrt{2}}{2}$. We have

$$\mathbf{Y}\mathbf{W}_{rot} = \begin{pmatrix} a & b \\ a & b \\ b & -a \\ b & -a \end{pmatrix} \quad \text{and} \quad \mathbf{Y}\mathbf{W}_{ref} = \begin{pmatrix} a & -b \\ a & -b \\ b & a \\ b & a \end{pmatrix},$$

   where $a^2 + b^2 = \frac{1}{2}$.

   Thus,

$$\mathbf{X}_{sp} - \mathbf{Y}\mathbf{W}_{rot} = \begin{pmatrix} -a & 1-b \\ -a & -b \\ \cos\alpha - b & a \\ \sin\alpha - b & a \end{pmatrix} \quad \text{and} \quad \mathbf{X}_{sp} - \mathbf{Y}\mathbf{W}_{ref} = \begin{pmatrix} -a & 1+b \\ -a & b \\ \cos\alpha - b & -a \\ \sin\alpha - b & -a \end{pmatrix}.$$

   Using the properties of trigonometric function, for the rotation case,

$$\min_{\mathbf{W}_{rot}} \|\mathbf{X}_{sp} - \mathbf{Y}\mathbf{W}_{rot}\|_{2,\infty} = \sqrt{\frac{1}{2} - \cos\alpha}$$

which occurs when the first and the third row of $\mathbf{X}_{sp} - \mathbf{Y}\mathbf{W}_{rot}$ have the same norm value.
For the reflection case, we have

$$\min_{\mathbf{W}_{ref}} \|\mathbf{X}_{sp} - \mathbf{Y}\mathbf{W}_{ref}\|_{2,\infty} = \sqrt{\frac{1}{2} + \sin\alpha}$$

which occurs when the first and the fourth row of $\mathbf{X}_{sp} - \mathbf{Y}\mathbf{W}_{rot}$ have the same norm value. Finally we choose the minimal between these two value which is $\sqrt{\frac{1}{2} - \cos\alpha}$.

Plug into the corresponding $\|\sin\Theta(\mathbf{X}_{sp}, \mathbf{Y})\|_2$, yields

$$\begin{aligned}
\|\mathbf{X}_{sp} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty} &= \sqrt{\frac{1}{2} - \cos\alpha} && (8)\\
&= \sqrt{\frac{1}{2} + \frac{|\cos\alpha - \sin\alpha|}{2} - \frac{|\cos\alpha + \sin\alpha|}{2}} && (9)\\
&= \sqrt{\frac{\sqrt{2}}{2}\|\sin\Theta(\mathbf{X}_{sp}, \mathbf{Y})\|_2 + \frac{1}{2} - \frac{\sqrt{2}}{2}\sqrt{1 - \|\sin\Theta(\mathbf{X}_{sp}, \mathbf{Y})\|_2^2}} \;. && (10)
\end{aligned}$$

2. $\alpha \in [\frac{3\pi}{2}, \frac{7\pi}{4}]$
By the properties of trigonometric functions, $|\sin\alpha + \frac{\sqrt{2}}{2}| \leq |\cos\alpha - \frac{\sqrt{2}}{2}| \leq 1 - \frac{\sqrt{2}}{2} \leq \frac{\sqrt{2}}{2} \leq |\cos\alpha + \frac{\sqrt{2}}{2}| \leq 1 \leq |\sin\alpha - \frac{\sqrt{2}}{2}| \leq 1 + \frac{\sqrt{2}}{2}$.
For the rotation case,

$$\min_{\mathbf{W}_{rot}} \|\mathbf{X}_{sp} - \mathbf{Y}\mathbf{W}_{rot}\|_{2,\infty} = \sqrt{\frac{1}{2} - \sin\alpha}$$

which occurs when the first and the fourth row of $\mathbf{X}_{sp} - \mathbf{Y}\mathbf{W}_{rot}$ have the same norm value.
For the reflection case, we have

$$\min_{\mathbf{W}_{ref}} \|\mathbf{X}_{sp} - \mathbf{Y}\mathbf{W}_{ref}\|_{2,\infty} = \sqrt{\frac{1}{2} + \cos\alpha}$$

which occur when the first and the third row of $\mathbf{X}_{sp} - \mathbf{Y}\mathbf{W}_{ref}$ have the same norm value. Finally, we choose the minimal between these two value which is $\sqrt{\frac{1}{2} + \cos\alpha}$.

Plug into the corresponding $\|\sin\Theta(\mathbf{X}_{sp}, \mathbf{Y})\|_2$, yields

$$\begin{aligned}
\|\mathbf{X}_{sp} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty} &= \sqrt{\frac{1}{2} + \cos\alpha} && (11)\\
&= \sqrt{\frac{1}{2} + \frac{|\cos\alpha - \sin\alpha|}{2} - \frac{|\cos\alpha + \sin\alpha|}{2}} && (12)\\
&= \sqrt{\frac{\sqrt{2}}{2}\|\sin\Theta(\mathbf{X}_{sp}, \mathbf{Y})\|_2 + \frac{1}{2} - \frac{\sqrt{2}}{2}\sqrt{1 - \|\sin\Theta(\mathbf{X}_{sp}, \mathbf{Y})\|_2^2}} \;. && (13)
\end{aligned}$$

Notice, when $\alpha \in [\frac{\pi}{2}, \frac{3\pi}{4}] \bigcup [\frac{3\pi}{2}, \frac{7\pi}{4}]$, $\|\sin\Theta(\mathbf{X}_{sp}, \mathbf{Y})\|_2 = \frac{|\cos\alpha - \sin\alpha|}{\sqrt{2}} \in [\frac{\sqrt{2}}{2}, 1]$. This shows for whatever value in $[\frac{\sqrt{2}}{2}, 1]$, we can find corresponding α satisfying it, thus existing a corresponding matrix $\mathbf{X}_{sp}$ satisfying it. In this way we construct the matrices satisfying $B(\mathbf{X}, \mathbf{Y})$.

□

4 $\mathbf{X} \in \mathbb{R}^{m\times 2}$ and $\mathbf{Y}$ is delocalized, Maximal Value of $\|\mathbf{X}-\mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$

We observe that as the dimension increases, the value of $\|\mathbf{X}-\mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$ tends to concentrate around regions where $\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2 \approx 1$. Given the complexity of capturing the full range of possible values for the $2,\infty$ norm (which would require a large number of simulations thus not necessarily possible to visualize), we shift our focus from finding an upper bound to identifying the largest possible value of $\|\mathbf{X}-\mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$. We know that this maximum occurs when $\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2 = 1$. Therefore, we propose a method to construct matrices where $\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2 = 1$, which allows us to approximate the potential maximal value of $\|\mathbf{X}-\mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$.

Consider $\mathbf{X},\mathbf{Y} \in \mathbb{R}^{m\times 2}$ where $m = 2d$ and $\mathbf{Y}$ is a delocalized matrix. We denote

$$\mathbf{X} = \begin{pmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_{\frac{m}{2}} & y_{\frac{m}{2}} \\ x_{\frac{m}{2}+1} & y_{\frac{m}{2}+1} \\ \vdots & \vdots \\ x_m & y_m \end{pmatrix} \quad \text{and} \quad \mathbf{Y} = \begin{pmatrix} \frac{1}{\sqrt{m}} & \frac{1}{\sqrt{m}} \\ \vdots & \vdots \\ \frac{1}{\sqrt{m}} & \frac{1}{\sqrt{m}} \\ \frac{1}{\sqrt{m}} & -\frac{1}{\sqrt{m}} \\ \vdots & \vdots \\ \frac{1}{\sqrt{m}} & -\frac{1}{\sqrt{m}} \end{pmatrix}.$$

We also denote

$$\Sigma_{i=1}^{\frac{m}{2}} x_i = m_1, \quad \Sigma_{i=\frac{m}{2}+1}^{m} x_i = m_2,$$

$$\Sigma_{i=1}^{\frac{m}{2}} y_i = m_3, \quad \Sigma_{i=\frac{m}{2}+1}^{m} y_i = m_4.$$

Consequently,

$$\mathbf{Y}^T\mathbf{X} = \begin{pmatrix} \frac{m_1+m_2}{\sqrt{m}} & \frac{m_3+m_4}{\sqrt{m}} \\ \frac{m_1-m_2}{\sqrt{m}} & \frac{m_3-m_4}{\sqrt{m}} \end{pmatrix}.$$

Taking the SVD of $\mathbf{Y}^T\mathbf{X}$, yields the diagonal matrix of singular values

$$\mathbf{\Sigma} = \begin{pmatrix} \frac{\sqrt{m_1^2+m_2^2+m_3^2+m_4^2-\sqrt{-4(m_2m_3-m_1m_4)^2+(m_1^2+m_2^2+m_3^2+m_4^2)^2}}}{\sqrt{m}} & 0 \\ 0 & \frac{\sqrt{m_1^2+m_2^2+m_3^2+m_4^2+\sqrt{-4(m_2m_3-m_1m_4)^2+(m_1^2+m_2^2+m_3^2+m_4^2)^2}}}{\sqrt{m}} \end{pmatrix}.$$

We denote $\sigma_2 = \mathbf{\Sigma}(1,1)$ and $\sigma_1 = \mathbf{\Sigma}(2,2)$ where $\sigma_1 \geq \sigma_2$. If $m_2m_3 - m_1m_4 = 0$, then $\sigma_2 = 0$, hence $\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2 = 1$. We turn our attention constructions satisfying $m_2m_3 - m_1m_4 = 0$.

Clearly, $m_2m_3 - m_1m_4 = 0$ if $m_2 = 0$ and $m_1 = 0$ or $m_3 = 0$ and $m_4 = 0$. In such settings, $\mathbf{X}$ is a very sparse matrix. We propose a special matrix below which has the largest candidate $\|\mathbf{X}-\mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$ value.

Claim 4.1 *If $m = 2d$, $d \geq 4$ is an integer, $\mathbf{X} \in \mathbb{O}^{m\times 2}$ and $\mathbf{Y}$ is a delocalized matrix, then*

$$\max_{\mathbf{X}} \|\mathbf{X}-\mathbf{Y}\mathbf{W}^*_{\mathbf{2},\infty}\|_{2,\infty} \geq \sqrt{\frac{1}{2}+\frac{2}{m}+\sqrt{\frac{2}{m}}}.$$

Proof: We construct the proof by showing there exist matrix $\mathbf{X}_{max}$ can achieve the equality we propose above. Such matrix is

$$\mathbf{X}_{max} = \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ \vdots & \vdots \\ 0 & \frac{\sqrt{2}}{2} \\ 0 & \frac{-\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & 0 \\ \frac{-\sqrt{2}}{2} & 0 \end{pmatrix}.$$

For $\|\mathbf{X}_{max} - \mathbf{Y}\mathbf{W}^*_{\mathbf{2},\infty}\|_{2,\infty}$, the first $m-4$ rows have the same l_2 norm $\sqrt{\frac{2}{m}}$ and the last 4 rows have the same l_2 norm $\sqrt{\frac{1}{2} + \frac{2}{m} + \sqrt{\frac{2}{m}}}$. This yields

$$\|\mathbf{X}_{max} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty} = \sqrt{\frac{1}{2} + \frac{2}{m} + \sqrt{\frac{2}{m}}} \quad \text{for} \quad m \geq 8.$$

□

By Lemma 1.4, we have $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_F\|_{2,\infty} \leq \sqrt{\frac{2}{m}} + \sqrt{\frac{m-2}{m}}$ which immediately serves as an upper bound for $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$. We denote the difference between the bound and $\|\mathbf{X}_{max} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|$ as

$$D(m) = \sqrt{\frac{2}{m}} + \sqrt{\frac{m-2}{m}} - \sqrt{\frac{1}{2} + \frac{2}{m} + \sqrt{\frac{2}{m}}}.$$

By direct calculation,

$$\lim_{m\to\infty} D(m) = 1 - \frac{\sqrt{2}}{2} \approx 0.29$$

This shows $\mathbf{W}^*_F$ is a good substitute for $\mathbf{W}^*_{2,\infty}$ at least considering the maximal value of $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{\mathbf{2},\infty}\|_{2,\infty}$. We list some conjectured largest $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{\mathbf{2},\infty}\|_{2,\infty}$ value below:

1. For 4×2 case, the conjectured largest value is obtained by the matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \\ \frac{\sqrt{2}}{2} & 0 \\ \frac{-\sqrt{2}}{2} & 0 \end{pmatrix}$, and the conjectured largest value is $\sqrt{\frac{1}{2} + \frac{\sqrt{2}}{2}}$.

2. For 8×2 case, the conjectured largest value is obtained by the matrix $\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \frac{\sqrt{2}}{2} \\ 0 & \frac{-\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & 0 \\ \frac{-\sqrt{2}}{2} & 0 \end{pmatrix}$, and the conjectured largest value is $\sqrt{\frac{5}{4}}$.

Combining the figures presented in Section 2 with the values of $\|\mathbf{X}_{\max} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$, we observe that while large values can occur in higher dimensions, the results are highly concentrated. Therefore, in Section 4, we shift our focus to these concentration inequalities.

5 Bounds in High Dimensions

Instead of the method for generating $\mathbf{X} \in \mathbb{R}^{m\times 2}$ in Section 2, we now generate the entries of $\mathbf{X}$ independently from $\mathcal{N}(0, \frac{1}{m})$. $\mathbf{Y}$ is a fixed $m \times 2$ delocalized matrix we mentioned above and $\mathbf{W}$ is a 2×2 orthogonal matrix. For $\mathbf{X}$, we denote two columns of $\mathbf{X}$ as $\vec{x}$ and $\vec{y}$. We denote the distribution of $\mathbf{X}$ as $\mathcal{S}$.

5.1 Upper bound for Expectation when Y is a delocalized matrix

Lemma 5.1 *Given fixed $\mathbf{Y} \in \mathbb{R}^{m\times 2}$, if we know $\mathbf{X} \sim \mathcal{S}$ where $\mathbf{X} \in \mathbb{R}^{m\times 2}$ and $\mathbf{W} \in \mathbb{O}^{2\times 2}$, then*

$$\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\min_{\mathbf{W}\in\mathbb{O}^{2\times 2}} \|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_{2,\infty}] \leq \min_{\mathbf{W}\in\mathbb{O}^{2\times 2}} \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_{2,\infty}].$$

Proof: Denote $f(\mathbf{X},\mathbf{W}) = \|\mathbf{X}-\mathbf{Y}\mathbf{W}\|_{2,\infty}$ and $\Phi(\mathbf{X}) = \min_{\mathbf{W}\in\mathbb{O}^{2\times 2}} \|\mathbf{X}-\mathbf{Y}\mathbf{W}\|_{2,\infty}$. By definition we have

$$0 \le \Phi(\mathbf{X}) \le f(\mathbf{X},\mathbf{W}) \quad \text{for any } \mathbf{W} \quad \text{with probablity 1.}$$

Taking the expectation with respect to $\mathbf{X}\sim\mathcal{S}$ we have

$$\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\Phi(\mathbf{X})] \le \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[f(\mathbf{X},\mathbf{W})] \quad \text{for any } \mathbf{W}.$$

Taking infimum on both sides yields

$$\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\min_{\mathbf{W}\in\mathbb{O}^{2\times 2}} \|\mathbf{X}-\mathbf{Y}\mathbf{W}\|_{2,\infty}] \le \min_{\mathbf{W}\in\mathbb{O}^{2\times 2}} \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\|\mathbf{X}-\mathbf{Y}\mathbf{W}\|_{2,\infty}].$$

□

Lemma 5.2 (Near-unit vector for iid Gaussian entry vector in high dimensions) *$\vec{x}$ and $\vec{y}$ are unit vector with high probability when m is large.*

Proof: See appendix. □

Lemma 5.3 (Near-orthogonality of Gaussian vectors in high dimensions) *$\vec{x}$ and $\vec{y}$ is orthogonal to each other with high probability when m is large.*

Proof: See appendix. □

These two lemmas allow us to obtain similar results, but using a simpler way to generate matrices.

Definition 5.1 (Sub-Gamma Random Variables, Section 2.4 in [BLM13]) *A real-valued centered random variable X is said to be sub-gamma on the right tail with variance factor v and scale parameter c if $\psi_X(\lambda) \le \frac{\lambda^2 v}{2(1-c\lambda)}$ for every λ such that $0<\lambda<1/c$. We denote the collection of such random variables by $\Gamma_+(v,c)$.*

Theorem 5.4 (Corollary 2.6 in [BLM13]) *Let $Z_1,\dots,Z_N$ be real-valued random variables belonging to $\Gamma_+(v,c)$. Then*

$$\mathbb{E}\max_{i=1,\dots,N} Z_i \le \sqrt{2v\log N} + c\log N.$$

Theorem 5.5 *Suppose $\mathbf{X}\in\mathbb{R}^{m\times 2}$ whose entries follow i.i.d. $\mathcal{N}(0,\frac{1}{m})$ and $\mathbf{Y}$ is a delocalized matrix, then*

$$\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\min_{\mathbf{W}\in\mathbb{O}^{2\times 2}} \|\mathbf{X}-\mathbf{Y}\mathbf{W}\|_{2,\infty}] \le \sqrt{\sqrt{\frac{8}{m^2}\log m + \frac{2}{m}\log m} + \frac{2}{m} + 2\sqrt{\frac{2}{m}}\sqrt{\sqrt{\frac{4}{m^2}\log 2m + \frac{2}{m}\log 2m} + \frac{1}{m}} + \frac{2}{m}}.$$

Proof: $\mathbf{W}$ can be rotation matrix or reflection matrix, if $\mathbf{Y}$ is a fixed delocalized matrix, by calculation we will have

$$\mathbf{Y}\mathbf{W}_{rot} = \begin{pmatrix} a & b \\ \vdots & \vdots \\ a & b \\ b & -a \\ \vdots & \vdots \\ b & -a \end{pmatrix} \quad \text{and} \quad \mathbf{Y}\mathbf{W}_{ref} = \begin{pmatrix} a & -b \\ \vdots & \vdots \\ a & -b \\ b & a \\ \vdots & \vdots \\ b & a \end{pmatrix}$$

where $a^2+b^2=\frac{2}{m}$. We denote $f_{rot}(\mathbf{X}) = \|\mathbf{X}-\mathbf{Y}\mathbf{W}_{rot}\|_{2,\infty}^2$ and $f_{ref}(\mathbf{X}) = \|\mathbf{X}-\mathbf{Y}\mathbf{W}_{ref}\|_{2,\infty}^2$. Then

$$\mathbb{E}[f_{rot}(\mathbf{X},a,b)] = \mathbb{E}[\max\left\{(x_1-a)^2+(y_1-b)^2,\dots,(x_m-b)^2+(y_m+a)^2\right\}],$$
$$\mathbb{E}[f_{ref}(\mathbf{X},a,b)] = \mathbb{E}[\max\left\{(x_1-a)^2+(y_1+b)^2,\dots,(x_m-b)^2+(y_m-a)^2\right\}].$$

By Lemma 5.1 we have

$$\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}\left[\min_{\mathbf{W}\in\mathbb{O}^{2\times 2}} \|\mathbf{X}-\mathbf{Y}\mathbf{W}\|_{2,\infty}\right] \le \min_{\mathbf{W}\in\mathbb{O}^{2\times 2}} \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\|\mathbf{X}-\mathbf{Y}\mathbf{W}\|_{2,\infty}].$$

Similarly we have

$$\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\min_{a^2+b^2=\frac{2}{m}} f_{rot}(\mathbf{X},a,b)] \le \min_{a^2+b^2=\frac{2}{m}} \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[f_{rot}(\mathbf{X},a,b)],$$

$$\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\min_{a^2+b^2=\frac{2}{m}} f_{ref}(\mathbf{X},a,b)] \le \min_{a^2+b^2=\frac{2}{m}} \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[f_{ref}(\mathbf{X},a,b)].$$

Given $\mathbf{X}$,

$$\min_{\mathbf{W}\in\mathcal{O}^{2\times 2}} \|\mathbf{X}-\mathbf{Y}\mathbf{W}\|_{2,\infty} = \min\left\{\min_{a^2+b^2=\frac{2}{m}} f_{rot}(\mathbf{X},a,b), \min_{a^2+b^2=\frac{2}{m}} f_{ref}(\mathbf{X},a,b)\right\}.$$

We will have

$$\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\min_{\mathbf{W}\in\mathcal{O}^{2\times 2}} \|\mathbf{X}-\mathbf{Y}\mathbf{W}\|_{2,\infty}] \le \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\min_{a^2+b^2=\frac{2}{m}} f_{rot}(\mathbf{X},a,b)],$$

$$\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\min_{\mathbf{W}\in\mathcal{O}^{2\times 2}} \|\mathbf{X}-\mathbf{Y}\mathbf{W}\|_{2,\infty}] \le \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\min_{a^2+b^2=\frac{2}{m}} f_{ref}(\mathbf{X},a,b)].$$

Then we continue to the expectation part.

We have $x_i, y_i \sim \mathcal{N}(0,\frac{1}{m})$ where $i\in\{1,2,\dots,m\}$ and we denote $L = \max\{|x_1|,|x_2|,\dots,|x_m|,|y_1|,|y_2|,\dots,|y_m|\}$, then for any a,b

$$\begin{aligned}
\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}\left[f_{rot}(\mathbf{X},a,b)\right] &= \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}\left[\max\left\{x_1^2+y_1^2+\frac{2}{m}-2ax_1-2by_1,\dots,x_m^2+y_m^2+\frac{2}{m}-2bx_m+2ay_m\right\}\right]\\
&\le \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}\left[\max\left\{x_1^2+y_1^2+\frac{2}{m}+2\left|ax_1\right|+2\left|by_1\right|,\dots,x_m^2+y_m^2+\frac{2}{m}+2\left|bx_m\right|+2\left|ay_m\right|\right\}\right]\\
&= \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}\left[\max\left\{x_1^2+y_1^2+2\left|ax_1\right|+2\left|by_1\right|,\dots,x_m^2+y_m^2+2\left|bx_m\right|+2\left|ay_m\right|\right\}\right]+\frac{2}{m}\\
&\le \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}\left[\max\left\{x_1^2+y_1^2+2(|a|+|b|)L,\dots,x_m^2+y_m^2+2(|a|+|b|)L\right\}\right]+\frac{2}{m}\\
&= \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}\left[\max\left\{x_1^2+y_1^2,\dots,x_m^2+y_m^2\right\}\right]+2(|a|+|b|)\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[L]+\frac{2}{m}\\
&= \underbrace{\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}\left[\max\left\{x_1^2+y_1^2,\dots,x_m^2+y_m^2\right\}\right]}_{(*)}+\underbrace{2(|a|+|b|)\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}\left[\max\left\{\left|x_1\right|,\dots,\left|x_m\right|,\left|y_1\right|,\dots,\left|y_m\right|\right\}\right]}_{(**)}+\frac{2}{m}
\end{aligned}$$

We construct our bounds following the definition of $\Gamma_+(v,c)$ in [BLM13], and use a similar method as the Example 2.7 in [BLM13].

1. For $(*)$, $X_i^* \sim \Gamma(1,\frac{2}{m})$ where $i=1,2,\dots,m$, then by definition $X_i^* \sim \Gamma_+(\frac{4}{m^2},\frac{2}{m})$.
   Let $M_m^* = \max\{X_1^*,X_2^*,\dots,X_m^*\}$, then

$$(*) = \mathbb{E}[M_m^*] \le \sqrt{\frac{8}{m^2}\log m}+\frac{2}{m}\log m+\frac{2}{m}.$$

2. For $(**)$, $X_i^{**} \sim \Gamma(\frac{1}{2},\frac{2}{m})$ where $i=1,2,\dots,2m$. By definition, we have $X_i^{**} \sim \Gamma_+(\frac{2}{m^2},\frac{2}{m})$. Let $M_{2m}^{**} = \max\{X_1^{**},X_2^{**},\dots,X_{2m}^{**}\}$, then

$$\mathbb{E}[M_{2m}^{**}] \le \sqrt{\frac{4}{m^2}\log 2m}+\frac{2}{m}\log 2m+\frac{1}{m},$$

$$(**) \le 2(|a|+|b|)\mathbb{E}[\sqrt{M_{2m}^{**}}].$$

Then we have

$$\min_{a^2+b^2=\frac{2}{m}} \mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[f_{rot}(\mathbf{X},a,b)] \le \sqrt{\frac{8}{m^2}\log m}+\frac{2}{m}\log m+\frac{2}{m}+2\sqrt{\frac{2}{m}}\sqrt{\sqrt{\frac{4}{m^2}\log 2m}+\frac{2}{m}\log 2m+\frac{1}{m}}+\frac{2}{m}.$$

For reflection function, the bound is the same. In conclusion, for delocalized $\mathbf{Y}$, we have

$$\mathbb{E}_{\mathbf{X}\sim\mathbb{S}}[\min_{\mathbf{W}\in\mathbb{O}^{2\times 2}}\|\mathbf{X}-\mathbf{YW}\|_{2,\infty}] \leq \sqrt{\sqrt{\frac{8}{m^2}\log m+\frac{2}{m}\log m+\frac{2}{m}}+2\sqrt{\frac{2}{m}}\sqrt{\sqrt{\frac{4}{m^2}\log 2m+\frac{2}{m}\log 2m+\frac{1}{m}}+\frac{2}{m}}},$$

which completes the proof. □

The following are some empirical results:

m (column number of $\mathbf{X}$)	Upper Bound in Theorem 5.5	Empirical Average of $\|\mathbf{X}-\mathbf{YW}^*_{2,\infty}\|_{2,\infty}$, entries of $\mathbf{X}$ generated from $\mathcal{N}(0,\frac{1}{m})$	Empirical Average of $\|\mathbf{X}-\mathbf{YW}^*_{2,\infty}\|_{2,\infty}$, entries of $\mathbf{X}$ generated from $\mathcal{N}(0,1)$ and then orthogonalize $\mathbf{X}$	Ratio of the third column to the second column
50	0.75	0.445	0.4488	0.593
100	0.55	0.348	0.35	0.633
200	0.406	0.266	0.268	0.655
300	0.338	0.223	0.2239	0.66
400	0.297	0.20	0.20	0.673
1000	0.196	0.137	0.14	0.699
2000	0.142	0.10	0.1005	0.704
5000	0.093	0.067	0.066	0.72
10000	0.067	0.049	0.049	0.731

Remark 2 *The ratio of the third column to the second column increases as m becomes larger.*

5.2 Symmetry in High Dimension

Since the entries of $\mathbf{X}$ and $\mathbf{Y}$ are drawn from $\mathcal{N}(0,\frac{1}{m})$, the squared ℓ_2 norm of each row of $\mathbf{X}-\mathbf{YW}_{\text{ref}}$ and $\mathbf{X}-\mathbf{YW}_{\text{rot}}$ follows the same distribution for fixed values of a and b. Specifically, for given a and b, the expression $(x-a)^2+(y-b)^2$ has the same distribution as $(x+a)^2+(y-b)^2$, where $x,y\sim\mathcal{N}(0,\frac{1}{m})$. This holds because both are shifted Gaussian distributions, and squaring them results in identical distributions.

Given this symmetry, we can visualize the problem. Regardless of whether $\mathbf{W}$ is a rotation or reflection matrix, each row of $\mathbf{X}-\mathbf{YW}$ follows the same distribution. Thus, the ℓ_2 norm of each row of $\mathbf{X}-\mathbf{YW}$, i.e., $\sqrt{(x_i-a)^2+(y_i-b)^2}$, can be interpreted as the distance from the point (x_i,y_i) to the point (a,b), where $a^2+b^2=\frac{2}{m}$. The pairs (a,b) lie on a circle, while (x_i,y_i), for $i=1,\ldots,m$, are drawn from $\mathcal{N}(0,\frac{1}{m})$.

We aim to find (a,b) that minimizes the maximum ℓ_2 norm of each row. With sufficiently many (x_i,y_i), the maximal distance $\sqrt{(x_i-a)^2+(y_i-b)^2}$ is expected to be similar for different (a,b). Our goal is to find the (a,b) that minimizes this maximum distance.

In the illustration below, the red points represent (a,b), and the blue points represent the (x_i,y_i) pairs (a total of m blue points). For each (a,b) on the red circle, there is a corresponding blue point that has the maximal distance from it. Therefore, finding $\min_{a,b}\|\mathbf{X}-\mathbf{YW}\|_{2,\infty}$ is equivalent to finding the point (a,b) on the red circle that minimizes the maximum distance to the blue points.

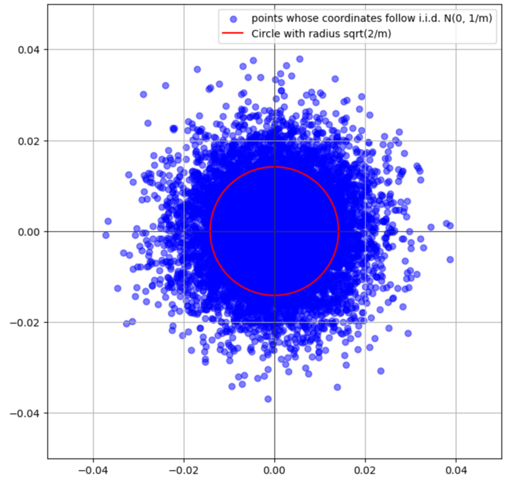


Figure 1: In this figure, m is chosen to be 10000.

5.3 High Probability Bound

Theorem 5.6 (Tail Bound for Sub-gamma Random Variable, Section 2.4, Characterization in [BLM13]) *Suppose that X is centered sub-gamma belongs to $\Gamma(\nu, c)$. Then*

$$\mathbb{P}[X > \sqrt{2\nu t} + ct] \leq \exp(-t).$$

Lemma 5.7 *For $X, Y \sim \mathcal{N}(0, \frac{1}{m})$, $Z = X^2 + Y^2$, we will have*

$$\mathbb{P}\{Z - \frac{2}{m} \geq \sqrt{\frac{8t}{m^2} + \frac{2t}{m}}\} \leq \exp(-t).$$

Proof: By definition $Z - \frac{2}{m} \sim \Gamma(\frac{4}{m^2}, \frac{2}{m})$. The rest is a direct use of Theorem 5.6. □

Lemma 5.8 *For any X and $Y \sim \mathcal{N}(0, \frac{1}{m})$, for any a, b where $a^2 + b^2 = \frac{2}{m}$, we have*

$$(X - a)^2 + (Y - b)^2 \leq X^2 + Y^2 + 2\sqrt{\frac{2}{m}(X^2 + Y^2)} + \frac{2}{m}.$$

Proof: Let $a = \sqrt{\frac{2}{m}} \cos\alpha$ and $b = \sqrt{\frac{2}{m}} \sin\alpha$.

$$\begin{aligned}(X - a)^2 + (Y - b)^2 &= X^2 + Y^2 + \frac{2}{m} - 2\sqrt{\frac{2}{m}} X \cos\alpha - 2\sqrt{\frac{2}{m}} Y \sin\alpha \\ &\leq X^2 + Y^2 + \frac{2}{m} + 2\sqrt{\frac{2}{m}(X^2 + Y^2)} \cos(\alpha + \Phi) \\ &\leq X^2 + Y^2 + \frac{2}{m} + 2\sqrt{\frac{2}{m}(X^2 + Y^2)}.\end{aligned}$$

□

Theorem 5.9 *For $\mathbf{X} \in \mathbb{R}^{m\times 2}$ whose entries follow $\mathcal{N}(0, \frac{1}{m})$, $\mathbf{Y} \in \mathbb{R}^{m\times 2}$ which is a delocalized matrix, with probability at least $1 - \exp(-mt)$,*

$$\min_{\mathbf{W}\in\mathbb{O}^{2\times 2}} \|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_{2,\infty} \leq \sqrt{\frac{2}{m}(2 + \sqrt{2t} + t + 2\sqrt{1 + \sqrt{2t} + t})}.$$

Proof: By Lemma 5.7 and Lemma 5.8, for X and $Y \sim \mathcal{N}(0, \frac{1}{m})$, we have, with probability at least $1 - \exp(-t)$,

$$(X-a)^2 + (Y-b)^2 \leq (\frac{2}{m} + \sqrt{\frac{8t}{m^2}} + \frac{2t}{m}) + \frac{2}{m} + 2\sqrt{\frac{2}{m}(\frac{2}{m} + \sqrt{\frac{8t}{m^2}} + \frac{2t}{m})}. \tag{14}$$

Since we have x_i and $y_i \sim \mathcal{N}(0, \frac{1}{m})$, then with probability less than $\exp(-tm)$, for all $i \in \{1, \ldots, m\}$, we have

$$(x_i-a)^2 + (y_i-b)^2 > (\frac{2}{m} + \sqrt{\frac{8t}{m^2}} + \frac{2t}{m}) + \frac{2}{m} + 2\sqrt{\frac{2}{m}(\frac{2}{m} + \sqrt{\frac{8t}{m^2}} + \frac{2t}{m})} \quad \text{for all } i. \tag{15}$$

It yields with probability at least $1 - \exp(-tm)$, there exists one pair of (x_j, y_j), such that

$$(x_j-a)^2 + (y_j-b)^2 \leq (\frac{2}{m} + \sqrt{\frac{8t}{m^2}} + \frac{2t}{m}) + \frac{2}{m} + 2\sqrt{\frac{2}{m}(\frac{2}{m} + \sqrt{\frac{8t}{m^2}} + \frac{2t}{m})}. \tag{16}$$

This is the same as with probability at least $1 - \exp(-mt)$,

$$\min_{\mathbf{W} \in \mathbb{O}^{2\times 2}} \|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_{2,\infty} \leq \sqrt{(\frac{2}{m} + \sqrt{\frac{8t}{m^2}} + \frac{2t}{m}) + \frac{2}{m} + 2\sqrt{\frac{2}{m}(\frac{2}{m} + \sqrt{\frac{8t}{m^2}} + \frac{2t}{m})}} \tag{17}$$

$$= \sqrt{\frac{2}{m}(2 + \sqrt{2t} + t + 2\sqrt{1 + \sqrt{2t} + t})} \tag{18}$$

which completes the proof. □

The following are some empirical results when we choose $1 - \exp(-t)$ to be 99% and $1 - \exp(-mt)$ is around 1. For the third column, we use the results we get previously (See Claim 4.1), and it shows the high probability bounds explain the concentration we want.

m (column number of $\mathbf{X}$)	High Probability Bound	Conjecture Maximal Value in Claim 4.1	Empirical Maximal of $\|\mathbf{X} - \mathbf{Y}\mathbf{W}^*_{2,\infty}\|_{2,\infty}$, entries of $\mathbf{X}$ generated from $\mathcal{N}(0, \frac{1}{m})$
50	0.787	0.860	0.582 (2000 iterations)
100	0.557	0.813	0.441 (1000 iterations)
200	0.393	0.781	0.310 (1000 iterations)
500	0.249	0.753	0.208 (600 iterations)
1000	0.176	0.739	0.158 (500 iterations)
2000	0.124	0.730	0.110 (200 iterations)
5000	0.078	0.721	0.072 (100 iterations)
10000	0.056	0.717	0.055 (100 iterations)

5.4 Upper Bound for Expectation, General Case

If $\mathbf{X}$ and $\mathbf{Y}$ are generated in the same manner, with each entry drawn independently from $\mathcal{N}(0, \frac{1}{m})$, then the squared l_2 norm of each row of $\mathbf{X} - \mathbf{Y}\mathbf{W}$ follows the same distribution, whether $\mathbf{W}$ is a rotation or reflection matrix, as discussed earlier in section 5.1.

Theorem 5.10 *If we have $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m\times n}$ and the entries of $\mathbf{X}$ and $\mathbf{Y}$ are generated independently from $\mathcal{N}(0, \frac{1}{m})$. Then*

$$\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\min_{\mathbf{W}\in\mathbb{O}^{2\times 2}} \|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_{2,\infty}] \leq \sqrt{\frac{4}{m}\sqrt{2\log m} + \frac{4}{m}\log m + \frac{4}{m}}.$$

Proof: Without loss of generality, we let $\mathbf{W}$ be a rotation matrix, i.e.,

$$\mathbf{W} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

We write

$$\mathbf{X} = \begin{pmatrix} a_1 & b_1 \\ \vdots & \vdots \\ a_m & b_m \end{pmatrix} \quad \text{and} \quad \mathbf{Y} = \begin{pmatrix} c_1 & d_1 \\ \vdots & \vdots \\ c_m & d_m \end{pmatrix}.$$

where $a_1, \ldots, a_m, b_1, \ldots, b_m, c_1, \ldots, c_m, d_1, \ldots, d_m \sim \mathcal{N}(0, \frac{1}{m})$. By calculation,

$$\mathbf{YW} = \begin{pmatrix} c_1 \cos\theta + d_1 \sin\theta & -c_1 \sin\theta + d_1 \cos\theta \\ \vdots & \vdots \\ c_m \cos\theta + d_m \sin\theta & -c_m \sin\theta + d_m \cos\theta \end{pmatrix}.$$

Since $c_i, d_i \sim \mathcal{N}\left(0, \frac{1}{m}\right)$, and linear combinations of independent normal random variables are still normally distributed, each entry of $\mathbf{YW}$ also follows a Gaussian distribution with mean 0 and variance $\frac{1}{m}$. Therefore, the entries of $\mathbf{YW}$ are i.i.d. $\mathcal{N}\left(0, \frac{1}{m}\right)$. If $\mathbf{X}_{ij} \sim \mathcal{N}\left(0, \frac{1}{m}\right)$ and $\mathbf{YW}_{ij} \sim \mathcal{N}\left(0, \frac{1}{m}\right)$, then $(\mathbf{X} - \mathbf{YW})_{ij} \sim \mathcal{N}\left(0, \frac{1}{m} + \frac{1}{m}\right) = \mathcal{N}\left(0, \frac{2}{m}\right)$.

By previous calculation, the squared l_2 norm (denote as $z_1, \ldots, z_m$) of each row of $\mathbf{X} - \mathbf{YW}$ follows $\Gamma(1, \frac{4}{m})$. This is the same as $\Gamma_+(\frac{16}{m^2}, \frac{4}{m})$. By Theorem 5.4, we have

$$\mathbb{E}[\max\{z_1, \ldots, z_m\}] \leq \frac{4}{m}\sqrt{2\log m} + \frac{4}{m}\log m + \frac{4}{m}$$

Then we have,

$$\mathbb{E}_{\mathbf{X}\sim\mathcal{S}}[\min_{\mathbf{W}\in\mathbb{O}^{2\times 2}} \|\mathbf{X} - \mathbf{YW}\|_{2,\infty}] \leq \sqrt{\frac{4}{m}\sqrt{2\log m} + \frac{4}{m}\log m + \frac{4}{m}}. \tag{19}$$

□

The following are some empirical results:

m (column number of X)	Upper Bound of Expectation of $\|\mathbf{X} - \mathbf{YW}^*_{2,\infty}\|_{2,\infty}$, entries of $\mathbf{X}$ and $\mathbf{Y}$ generated from $\mathcal{N}\left(0, \frac{1}{m}\right)$ (12)	Empirical Average of $\|\mathbf{X} - \mathbf{YW}^*_{2,\infty}\|_{2,\infty}$, entries of $\mathbf{X}$ and $\mathbf{Y}$ generated from $\mathcal{N}\left(0, \frac{1}{m}\right)$	ratio of the third column to the second column
50	0.785	0.471	0.6
100	0.588	0.371	0.63
200	0.437	0.285	0.652
300	0.366	0.243	0.663
400	0.323	0.218	0.675
1000	0.215	0.15	0.698
2000	0.158	0.111	0.703
5000	0.104	0.075	0.721
10000	0.076	0.056	0.789

6 Future Work

6.1 Gradient Descent

In this report, we do not delve into the details of the general case where $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m\times n}$ with $m \geq n$, as this requires more complex algorithms. However, gradient descent may still work well for this scenario with certain modifications.

Consider the case when $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m\times n}$ and $\mathbf{W} \in \mathbb{R}^{n\times n}$. In this case, we have n^2 parameters, corresponding to the entries of $\mathbf{W}$, and we define $\mathbf{Z} = \mathbf{X} - \mathbf{YW}$. Let $f_i(w_{11}, w_{12}, \ldots, w_{nn})$ represent the squared ℓ_2 norm of each row, where $i \in \{1, 2, \ldots, m\}$.

The loss function is given by:

$$L(w_{11}, w_{12}, \ldots, w_{nn}) = \max\{f_1, f_2, \ldots, f_m\}.$$

Since each f_i is a convex function, the loss function $L(w_{11}, w_{12}, \ldots, w_{nn})$ is also convex. However, the challenge lies in the fact that there are constraints on $\mathbf{W}$ — specifically, preserving its orthonormality — which gradient descent cannot directly maintain. This might be solved by adding a retraction each step.

For a given row f_p, we have:

$$f_p = \sum_{j=1}^{n} \left(\mathbf{X}_{pj} - \sum_{i=1}^{n} \mathbf{Y}_{pi} w_{ij} \right)^2 .$$

Thus, the partial derivative of f_p with respect to w_{ij} is:

$$\frac{\partial f_p}{\partial w_{ij}} = -2\mathbf{Y}_{pj} \left(\mathbf{X}_{pj} - \sum_{i=1}^{n} \mathbf{Y}_{pi} w_{ij} \right) = -2\mathbf{Y}_{pi}\mathbf{Z}_{pj}.$$

This is equivalent to selecting the p-th row of $\mathbf{Z}$ and $\mathbf{Y}$, denoted as $\mathbf{z}_p$ and $\mathbf{y}_p$, respectively. The matrix entries of $-2\mathbf{y}_p^T \mathbf{z}_p$ correspond to the entries of $\frac{\partial f_p}{\partial w_{ij}}$.

Additionally, for the Hessian of f_p, with $\bar{\mathbf{W}}$ denoting the vectorized form of $\mathbf{W}$, we have:

$$\nabla^2 f_p(\bar{\mathbf{W}}) = \begin{pmatrix} \mathbf{Y}_p \mathbf{Y}_p^T & 0 & \cdots & 0 \\ 0 & \mathbf{Y}_p \mathbf{Y}_p^T & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Y}_p \mathbf{Y}_p^T \end{pmatrix},$$

which is a positive semidefinite matrix.

While the loss function is convex, the constraint that $\mathbf{W}$ lies on the Stiefel manifold means that the set of feasible solutions is not convex. This poses additional challenges for optimization.

6.2 Convex Combination

This is an aspect we do not explore in detail. Specifically, for matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{4\times 2}$, a single parameter θ can be used to represent both the rotation matrix $\mathbf{W}_{\text{rot}}$ and the reflection matrix $\mathbf{W}_{\text{ref}}$. We can then investigate how closely the parameter θ_F, representing the optimal rotation or reflection matrix $\mathbf{W}_F^*$ in terms of the Frobenius norm, compares with $\theta_{2,\infty}$, which represents the matrix $\mathbf{W}_{2,\infty}^*$ optimized for the $2,\infty$ norm.

Additionally, we can define a new objective function as follows. Let

$$L_F(\theta) = \|\mathbf{X} - \mathbf{Y}\mathbf{W}(\theta)\|_F^2,$$

$$L_{2,\infty}(\theta) = \|\mathbf{X} - \mathbf{Y}\mathbf{W}(\theta)\|_{2,\infty}^2.$$

We can then construct a new objective function as a convex combination of these two:

$$L_c(\theta) = \lambda L_F(\theta) + (1-\lambda) L_{2,\infty}(\theta) \quad \text{where} \quad \lambda \in [0,1].$$

Since $L_F(\theta)$ is smooth while $L_{2,\infty}(\theta)$ is not, by choosing θ properly, we can make the new objective function smoother. We also ask: as we adjust λ, will θ_F gradually converge to $\theta_{2,\infty}$?

A potential challenge arises from the fact that $\mathbf{W}_F^*$ may be a rotation matrix while $\mathbf{W}_{2,\infty}^*$ could be a reflection matrix (or vice versa). This complicates the process of measuring how close θ_F and $\theta_{2,\infty}$ truly are.

6.3 Difference between $\|\mathbf{X} - \mathbf{Y}\mathbf{W}_F^*\|_{2,\infty}$ and $\|\mathbf{X} - \mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$

In this direction, we present some empirical results for low-dimensional cases. Refer to the figures below, where the y-axis represents the difference $\|\mathbf{X} - \mathbf{Y}\mathbf{W}_F^*\|_{2,\infty} - \|\mathbf{X} - \mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$.

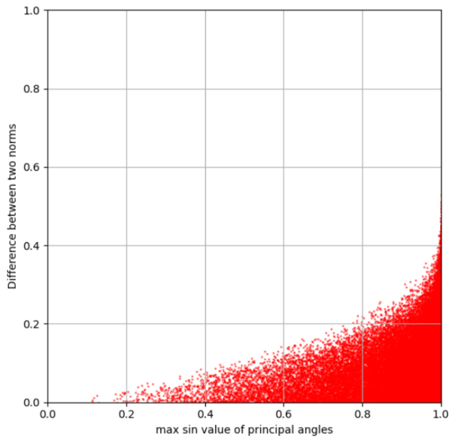


Figure 2: $\mathbf{X}, \mathbf{Y} \in \mathbb{O}^{4\times 2}$, 40000 points

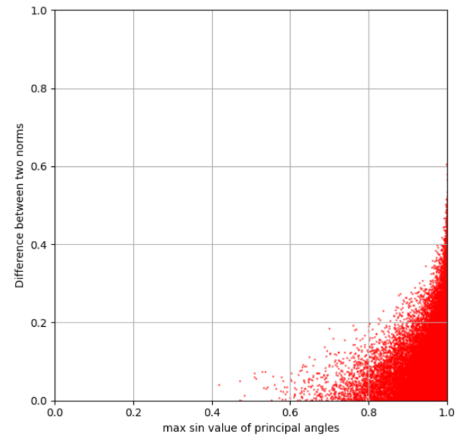


Figure 3: $\mathbf{X}, \mathbf{Y} \in \mathbb{O}^{8\times 2}$, 40000 points

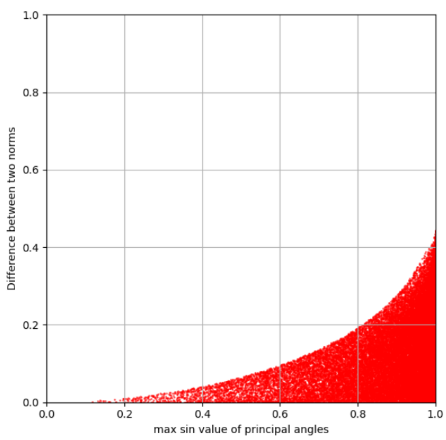


Figure 4: $\mathbf{X}, \mathbf{Y} \in \mathbb{O}^{4\times 2}$ and $\mathbf{Y}$ is delocalized, 40000 points

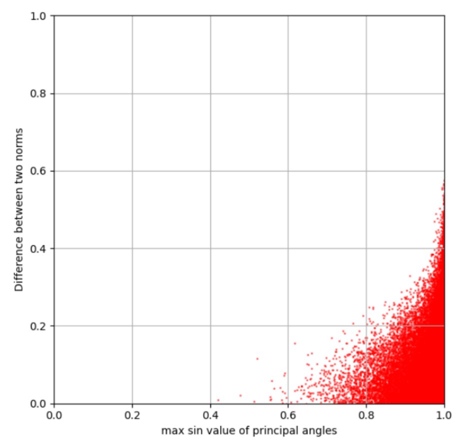


Figure 5: $\mathbf{X}, \mathbf{Y} \in \mathbb{O}^{8\times 2}$ and $\mathbf{Y}$ is delocalized, 40000 points

We observe that in higher dimensions, both $\|\mathbf{X}-\mathbf{Y}\mathbf{W}_F^*\|_{2,\infty}$ and $\|\mathbf{X}-\mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$ tend have $\|\sin\Theta(\mathbf{X},\mathbf{Y})\|_2$ closer to 1. For this reason, it makes sense to directly examine the difference $\|\mathbf{X}-\mathbf{Y}\mathbf{W}_F^*\|_{2,\infty} - \|\mathbf{X}-\mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$.

References

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Oxford University Press, 2013.

[Cap20] Joshua Cape. A note on the orthogonal procrustes problem and norm-dependent optimality. *The Electronic Journal of Linear Algebra*, 36:158–168, 2020.

Appendix

A Omitted Proof

Proof of Theorem 1.1:

Proof: We denote $\mathbf{W}_F^*$ to be $\operatorname{arginf}_{\mathbf{W}} \|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_F^2$, which is the same as $\operatorname{arginf}_{\mathbf{W}} \|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_F$, then we have

$$\begin{aligned}
\mathbf{W}_F^* &= \operatorname*{arginf}_{\mathbf{W}} \|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_F^2 && (20)\\
&= \operatorname*{arginf}_{\mathbf{W}} \langle \mathbf{X} - \mathbf{Y}\mathbf{W}, \mathbf{X} - \mathbf{Y}\mathbf{W} \rangle_F && (21)\\
&= \operatorname*{arginf}_{\mathbf{W}} \|\mathbf{X}\|_F^2 + \|\mathbf{Y}\mathbf{W}\|_F^2 - 2\langle \mathbf{X}, \mathbf{Y}\mathbf{W} \rangle_F && (22)
\end{aligned}$$

Since $\|\mathbf{Y}\mathbf{W}\|_F^2 = \operatorname{trace}\left(\mathbf{Y}\mathbf{W}(\mathbf{Y}\mathbf{W})^T\right) = \|\mathbf{Y}\|_F^2$ and $\mathbf{X}$ and $\mathbf{Y}$ are given, we can convert the problem above into

$$\begin{aligned}
\mathbf{W}_F^* &= \operatorname*{argsup}_{\mathbf{W}} \langle \mathbf{X}, \mathbf{Y}\mathbf{W} \rangle_F && (23)\\
&= \operatorname*{argsup}_{\mathbf{W}} \left\langle \mathbf{Y}^T\mathbf{X}, \mathbf{W} \right\rangle_F && (24)
\end{aligned}$$

By SVD, we can write $\mathbf{Y}^T\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ and we have

$$\begin{aligned}
\mathbf{W}_F^* &= \operatorname*{argsup}_{\mathbf{W}} \left\langle \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \mathbf{W} \right\rangle_F && (25)\\
&= \operatorname*{argsup}_{\mathbf{W}} \operatorname{trace}\left(\mathbf{W}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\right) && (26)\\
&= \operatorname*{argsup}_{\mathbf{W}} \operatorname{trace}\left(\mathbf{\Sigma}\mathbf{V}^T\mathbf{W}^T\mathbf{U}\right) && (27)\\
&= \operatorname*{argsup}_{\mathbf{W}} \langle \mathbf{\Sigma}, \mathbf{V}^T\mathbf{W}^T\mathbf{U} \rangle_F && (28)
\end{aligned}$$

Notice that

$$(\mathbf{V}^T\mathbf{W}^T\mathbf{U})^T\mathbf{V}^T\mathbf{W}^T\mathbf{U} = \mathbf{I}$$

Since $\mathbf{\Sigma}$ is a diagonal matrix, we have (28) is maximized when $\mathbf{V}^T\mathbf{W}^T\mathbf{U} = \mathbf{I}$, that is

$$\mathbf{W}_F^* = \mathbf{U}\mathbf{V}^T$$

□

Proof of Lemma 5.2:

Proof: Let $x_1, \ldots, x_m \sim \mathcal{N}(0, \frac{1}{m})$. Denote $Z = \Sigma_{i=1}^m x_i^2$, then $\mathbb{E}[Z] = 1$ and $Var[Z] = \frac{2}{m}$. By Chebyshev's inequality,

$$\mathbb{P}(|Z - 1| > \frac{k}{\sqrt{2m}}) \leq \frac{1}{k^2}.$$

The ℓ_2 norm of $\vec{x}$ is $\sqrt{Z}$. □

Proof of Lemma 5.3:

Proof: Let $x_1, \ldots, x_m, y_1, \ldots, y_m \sim \mathcal{N}(0, \frac{1}{m})$. Denote $z_i = x_i y_i$ and $Z = \Sigma_{i=1}^m z_i$, then $\mathbb{E}[z_i] = 0$ and $Var[z_i] = \frac{1}{m^2}$. Thus $\mathbb{E}[Z] = 0$ and $Var[Z] = \frac{1}{m}$. By Chebyshev's inequality,

$$\mathbb{P}(|Z| > \frac{k}{\sqrt{m}}) \leq \frac{1}{k^2}.$$

□

B Python Code for Computing $\|\mathbf{X} - \mathbf{Y}\mathbf{W}_{2,\infty}^*\|_{2,\infty}$

```python
import numpy as np
from scipy.linalg import svd
import matplotlib.pyplot as plt

def gram_schmidt(X):
    Q, _ = np.linalg.qr(X)
    return Q

def sample_orthogonal_matrix(m, r):
    Z = np.random.normal(size=(m, r))
    U = gram_schmidt(Z)
    return U

def generate_gaussian_matrix(m):
    matrix = np.random.normal(0, 1/np.sqrt(m), (m, 2))
    return matrix

def principal_angles(U, V):
    U1, sigma, V1 = svd(np.dot(V.T, U))
    angles = np.arccos(np.clip(sigma, -1.0, 1.0))
    return angles

def sin_principal_angles(U, V):
    angles = principal_angles(U, V)
    sin_angles = np.sin(angles)
    return np.linalg.norm(np.diag(sin_angles), ord=2)

def compute_optimal_norm(X, Y):
    def l2_norm(row):
        return np.linalg.norm(row)

    def calculate_row_norms(theta, W_type):
        if W_type == 'rotation':
            W = np.array([[np.cos(theta), -np.sin(theta)], [np.sin(theta), np.cos(theta)]])
        else:
            W = np.array([[np.cos(theta), np.sin(theta)], [np.sin(theta), -np.cos(theta)]])

        row_norms = []
        for i in range(X.shape[0]):
            row = X[i] - Y[i] @ W
            row_norms.append(l2_norm(row))
        return row_norms

    theta_values = np.linspace(0, 2 * np.pi, 1000)

    f_max_values_rotation = [np.max(calculate_row_norms(theta, 'rotation')) for theta in theta_values]
    f_max_values_reflection = [np.max(calculate_row_norms(theta, 'reflection')) for theta in theta_values]

    min_f_max_value = float('inf')
    best_theta = None
    best_type = None

    for i, theta in enumerate(theta_values):
        min_f_max = min(f_max_values_rotation[i], f_max_values_reflection[i])
        if min_f_max < min_f_max_value:
            min_f_max_value = min_f_max
            best_theta = theta
```

```
            best_type = 'rotation' if f_max_values_rotation[i] < f_max_values_reflection[i] else 'reflection'

    if best_type == 'rotation':
        best_W = np.array([[np.cos(best_theta), -np.sin(best_theta)], [np.sin(best_theta), np.cos(best_theta)]])
    else:
        best_W = np.array([[np.cos(best_theta), np.sin(best_theta)], [np.sin(best_theta), -np.cos(best_theta)]])

    return min_f_max_value
```